GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: November 5, 2001, 22:35:55 ; Search time 1157.88 Seconds
(without alignments)
4947.343 Million cell updates/sec

Title: US-09-656-668-198
Perfect score: 606
Sequence: 1 tgagtttgccccttacccc.....aagcctgttcttgtctgcac 606

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 0

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	313	51.7	408	15	AI023799	AI023799 ox08d07.x
2	275	45.4	405	12	AA826307	AA826307 od71f05.s
c 3	268	44.2	473	23	AI683094	AI683094 tx01e12.x
c 4	204	33.7	414	18	AI307373	AI307373 tb26f01.x
c 5	77	12.7	398	103	AI870547	AI870547 wl47a04.x
6	48	7.9	269	123	AW979007	AW979007 EST391117
c 7	29	4.8	234	19	AI366910	AI366910 qv94g09.x
c 8	25	4.1	428	24	AI767719	AI767719 wh39a02.x
c 9	25	4.1	530	22	AI609595	AI609595 tw28d08.x
c 10	25	4.1	541	120	AW771618	AW771618 hn59c06.x
11	23	3.8	323	164	BE144594	BE144594 MR0-HT016
12	23	3.8	390	256	B03713	B03713 cSRL-185e7-
c 13	23	3.8	528	226	AQ282338	AQ282338 RPCI11-79
14	21	3.5	422	188	R77087	R77087 yi64g04.r1
c 15	21	3.5	1018	105	AL513616	AL513616 AL513616
16	20	3.3	333	229	AQ491345	AQ491345 RPCI-11-2
17	20	3.3	525	165	BE237302	BE237302 146445 MA
c 18	20	3.3	528	232	AQ682294	AQ682294 HS_5500_B
c 19	20	3.3	582	148	BF396078	BF396078 UI-R-CA1-
20	20	3.3	630	231	AQ651011	AQ651011 Sheared D
c 21	20	3.3	630	238	AZ087326	AZ087326 RPCI-23-6
c 22	20	3.3	644	231	AQ638970	AQ638970 927P1-15A
23	20	3.3	665	248	AZ711584	AZ711584 RPCI-24-8
24	20	3.3	735	221	CNS04UDW	AL307661 Tetraodon
25	20	3.3	1094	222	CNS05BC4	AL329629 Tetraodon
26	20	3.3	1417	137	BE614327	BE614327 601504388
27	19	3.1	168	7	AA399547	AA399547 zt88b12.s
c 28	19	3.1	198	6	AA397582	AA397582 zt88b12.r
c 29	19	3.1	206	169	BF750054	BF750054 CM3-BN038
30	19	3.1	229	160	BB555515	BB555515 BB555515
c 31	19	3.1	249	115	AW353831	AW353831 33434 MAR
c 32	19	3.1	250	145	BF171105	BF171105 PCL2071 M
c 33	19	3.1	294	188	T33881	T33881 EST59687 Hu
c 34	19	3.1	299	191	Z45245	Z45245 HSC2KE081 n
35	19	3.1	308	6	AA363307	AA363307 EST73150
36	19	3.1	319	25	AV170290	AV170290 AV170290
37	19	3.1	334	188	T05488	T05488 EST03377 Fe
c 38	19	3.1	334	191	Z24994	Z24994 HSB81G012 S
c 39	19	3.1	336	191	Z25314	Z25314 HSBB9A082 S
c 40	19	3.1	351	150	BF557133	BF557133 UI-R-E1-g
41	19	3.1	390	224	AQ105014	AQ105014 HS_2168_B
c 42	19	3.1	415	17	AI188548	AI188548 qd09g07.x
c 43	19	3.1	419	188	R84776	R84776 yt65h10.r1
c 44	19	3.1	428	189	T87433	T87433 yd89f06.r1
45	19	3.1	443	239	AZ149979	AZ149979 SP_0031_B

ALIGNMENTS

RESULT 1
AI023799/c
LOCUS AI023799 408 bp mRNA EST 28-AUG-1998
DEFINITION ox08d07.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:1655725 3' similar to contains MER20.t2 MER20 repetitive element ;, mRNA sequence.
ACCESSION AI023799
VERSION AI023799.1 GI:3238843
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 408)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 371 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 375.
FEATURES Location/Qualifiers
source 1. .408
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1655725"
/clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen 1NFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer {5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTT 3'}, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 104 a 82 c 93 g 129 t
ORIGIN

Query Match 51.7%; Score 313; DB 15; Length 408;
Best Local Similarity 99.7%; Pred. No. 1.1e-158;
Matches 363; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    200 tgagaaaacctaccgcaggatcttactgggcttcataggtaagcttgcctttgttctggc 259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    408 TGAGAAAACCTACCGCAGGATCTTACTGGGCTTCATAGGTAAGCTTGCCTTTGTTCTGGC 349

Qy    260 ttctgtagatatataaaataaagacactgcccagtccctccctcaacgtcccgagccagg 319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    348 TTCTGTAGATATATAAAATAAAGACACTGCCCAGTCCCTCCCTCAACGTCCCGAGCCAGG 289

Qy    320 gctcaaggcaattccaataacagtagaatgaacactaaatattgatttcaaaatctcagc 379
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    288 GCTCAAGGCAATTCCAATAACAGTAGAATGAACCCTAAATATTGATTTCAAAATCTCAGC 229

Qy    380 aactagaagaatgaccaaccatcctggttggcctgggactgtcctagttttagcattgaa 439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    228 AACTAGAAGAATGACCAACCATCCTGGTTGGCCTGGGACTGTCCTAGTTTTAGCATTGAA 169

Qy    440 agtttcaggttccaggaaagccctcaggcctgggctgctggtcaccctagcagctgaggg 499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    168 AGTTTCAGGTTCCAGGAAAGCCCTCAGGCCTGGGCTGCTGGTCACCCTAGCAGCTGAGGG 109

Qy    500 actcttcaatacagaattagtctttgtgcactggagatgaatatactttaatttgtaaca 559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    108 ACTCTTCAATACAGAATTAGTCTTTGTGCACTGGAGATGAATATACTTTAATTTGTAACA 49

Qy    560 tgtg 563
          ||||
Db     48 TGTG 45
```

RESULT 2
AA826307
LOCUS AA826307 405 bp mRNA EST 07-APR-1998
DEFINITION od71f05.s1 NCI_CGAP_Ov2 Homo sapiens cDNA clone IMAGE:1373409, mRNA sequence.
ACCESSION AA826307
VERSION AA826307.1 GI:2899619
KEYWORDS EST.

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 500   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .405
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1373409"
                     /clone_lib="NCI_CGAP_Ov2"
                     /sex="female"
                     /tissue_type="ovary"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; mRNA made from invasive ovarian
                     tumor, cDNA made by oligo-dT priming.  Non-directionally
                     cloned. Size-selected on agarose gel, average insert  size
                     600 bp. Reference:  Krizman et al. (1996) Cancer Research
                     56:5380-5383."
BASE COUNT      118 a    102 c     79 g    106 t
ORIGIN
```

```
Query Match           45.4%;  Score 275;  DB 12;  Length 405;
Best Local Similarity  99.7%;  Pred. No. 4.5e-138;
Matches  325;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    272 ataaaataaagacactgcccagtccctccctcaacgtcccgagccagggctcaaggcaat 331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9 ATAAAATAAAGACACTGCCCAGTCCCTCCCTCAACGTCCCGAGCCAGGGCTCAAGGCAAT 68

Qy    332 tccaataacagtagaatgaacactaaatattgatttcaaaatctcagcaactagaagaat 391
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69 TCCAATAACAGTAGAATGAACACTAAATATTGATTTCAAAATCTCAGCAACTAGAAGAAT 128

Qy    392 gaccaaccatcctggttggcctgggactgtcctagttttagcattgaaagtttcaggttc 451
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    129 GACCAACCATCCTGGTTGGCCTGGGACTGTCCTAGTTTTAGCATTGAAAGTTTCAGGTTC 188

Qy    452 caggaaagccctcaggcctgggctgctggtcaccctagcagctgagggactcttcaatac 511
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    189 CAGGAAAGCCCTCAGGCCTGGGCTGCTGGTCACCCTAGCAGCTGAGGGACTCTTCAATAC 248

Qy    512 agaattagtctttgtgcactggagatgaatatactttaatttgtaacatgtgaaaacatc 571
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    249 AGAATTAGTCTTTGCGCACTGGAGATGAATATACTTTAATTTGTAACATGTGAAAACATC 308

Qy    572 tataaacatctactgaagcctgttct 597
          |||||||||||||||||||||||||||
Db    309 TATAAACATCTACTGAAGCCTGTTCT 334
```

```
RESULT   3
AI683094/c
LOCUS       AI683094      473 bp    mRNA            EST       16-DEC-1999
DEFINITION  tx01el2.x1 NCI_CGAP_Ut4 Homo sapiens cDNA clone IMAGE:2267950 3',
            mRNA sequence
ACCESSION   AI683094
VERSION     AI683094.1  GI:4893286
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 473)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 2912   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 405.
FEATURES             Location/Qualifiers
     source          1. .473
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2267950"
                     /clone_lib="NCI_CGAP_Ut4"
                     /tissue_type="serous papillary carcinoma, high grade, 2
                     pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.48 kb. Life Technologies catalog #:
                     11542-016"
BASE COUNT      142 a     88 c     95 g    148 t
ORIGIN
```

```
Query Match           44.2%;  Score 268;  DB 23;  Length 473;
Best Local Similarity  100.0%;  Pred. No. 2.9e-134;
Matches  268;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    330 attccaataacagtagaatgaacactaaatattgatttcaaaatctcagcaactagaaga 389
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    473 ATTCCAATAACAGTAGAATGAACACTAAATATTGATTTCAAAATCTCAGCAACTAGAAGA 414

Qy    390 atgaccaaccatcctggttggcctgggactgtcctagttttagcattgaaagtttcaggt 449
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    413 ATGACCAACCATCCTGGTTGGCCTGGGACTGTCCTAGTTTTAGCATTGAAAGTTTCAGGT 354

Qy    450 tccaggaaagccctcaggcctgggctgctggtcaccctagcagctgagggactcttcaat 509
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    353 TCCAGGAAAGCCCTCAGGCCTGGGCTGCTGGTCACCCTAGCAGCTGAGGGACTCTTCAAT 294

Qy    510 acagaattagtctttgtgcactggagatgaatatactttaatttgtaacatgtgaaaaca 569
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    293 ACAGAATTAGTCTTTGTGCACTGGAGATGAATATACTTTAATTTGTAACATGTGAAAACA 234

Qy    570 tctataaacatctactgaagcctgttct 597
          |||||||||||||||||||||||||||||
Db    233 TCTATAAACATCTACTGAAGCCTGTTCT 206
```

```
RESULT   4
AI307373/c
LOCUS       AI307373      414 bp    mRNA            EST       08-APR-1999
DEFINITION  tb26f01.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2055481 3',
            mRNA sequence.
ACCESSION   AI307373
```

```
VERSION     AI307373.1  GI:4002098
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 414)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 592   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .414
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2055481"
                     /clone_lib="NCI_CGAP_Kid12"
                     /tissue_type="2 pooled tumors (clear cell type)"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid5 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1323912-1325831, 1471368-1472903 and
                     1492104-1493255). Subtraction by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT      126 a     80 c     85 g    123 t
ORIGIN
```

```
  Query Match           33.7%;  Score 204;  DB 18;  Length 414;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-99;
  Matches  204;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    394 ccaaccatcctggttggcctgggactgtcctagttttagcattgaaagtttcaggttcca 453
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    414 CCAACCATCCTGGTTGGCCTGGGACTGTCCTAGTTTTAGCATTGAAAGTTTCAGGTTCCA 355

Qy    454 ggaaagccctcaggcctgggctgctggtcaccctagcagctgagggactcttcaatacag 513
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    354 GGAAAGCCCTCAGGCCTGGGCTGCTGGTCACCCTAGCAGCTGAGGGACTCTTCAATACAG 295

Qy    514 aattagtctttgtgcactggagatgaatatactttaatttgtaacatgtgaaaacatcta 573
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    294 AATTAGTCTTTGTGCACTGGAGATGAATATACTTTAATTTGTAACATGTGAAAACATCTA 235

Qy    574 taaacatctactgaagcctgttct 597
          |||||||||||||||||||||||
Db    234 TAAACATCTACTGAAGCCTGTTCT 211
```

```
RESULT   5
AI870547/c
LOCUS       AI870547      398 bp    mRNA            EST       07-MAR-2000
DEFINITION  wl47a04.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2428014 3',
            mRNA sequence.
ACCESSION   AI870547
VERSION     AI870547.1  GI:5544515
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 398)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1826   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 373.
FEATURES             Location/Qualifiers
     source          1. .398
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2428014"
                     /clone_lib="NCI_CGAP_Ut1"
                     /tissue_type="well-differentiated endometrial
                     adenocarcinoma, 7 pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Life Technologies catalog #:
                     11538-014"
BASE COUNT      130 a     56 c     68 g    144 t
ORIGIN
```

```
  Query Match           12.7%;  Score 77;  DB 103;  Length 398;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-30;
  Matches   77;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    521 ctttgtgcactggagatgaatatactttaatttgtaacatgtgaaaacatctataaacat 580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    398 CTTTGTGCACTGGAGATGAATATACTTTAATTTGTAACATGTGAAAACATCTATAAACAT 339

Qy    581 ctactgaagcctgttct 597
          |||||||||||||||||
Db    338 CTACTGAAGCCTGTTCT 322
```

```
RESULT   6
AW979007
LOCUS       AW979007      269 bp    mRNA            EST       02-JUN-2000
DEFINITION  EST391117 MAGE resequences, MAGP Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW979007
VERSION     AW979007.1  GI:8170290
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 269)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
```

```
          9712 Medical Center Dr., Rockville, MD 20850, USA
          Tel: 301 838 3528
          Fax: 301 838 0208
          Email: johnq@tigr.org
          Plate: 401
          Seq primer: Forward.
FEATURES             Location/Qualifiers
     source          1. .269
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGP"
                     /note="Vector: pBluescriptSKm"
BASE COUNT       82 a     69 c     50 g     55 t     13 others
ORIGIN

  Query Match             7.9%;  Score 48;  DB 123;  Length 269;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-15;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     272 ataaaataaagacactgcccagtccctccctcaacgtcccgagccagg 319
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db     113 ATAAAATAAAGACACTGCCCAGTCCCTCCCTCAACGTCCCGAGCCAGG 160


RESULT   7
AI366910/c
LOCUS       AI366910      234 bp    mRNA            EST       08-JAN-1999
DEFINITION  qv94g09.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:1989280 3',
            mRNA sequence.
ACCESSION   AI366910
VERSION     AI366910.1  GI:4136655
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 234)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 226.
FEATURES             Location/Qualifiers
     source          1. .234
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1989280"
                     /clone_lib="NCI_CGAP_Ut2"
                     /tissue_type="moderately-differentiated endometrial
                     adenocarcinoma, 3 pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.85 kb. Life Technologies catalog #:
                     11539-012"
BASE COUNT       54 a     42 c     38 g     100 t
ORIGIN

  Query Match             4.8%;  Score 29;  DB 19;  Length 234;
  Best Local Similarity   100.0%;  Pred. No. 0.00016;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     550 atttgtaacatgtgaaaacatctataaac 578
           ||||||||||||||||||||||||||||||
Db     122 ATTTGTAACATGTGAAAACATCTATAAAC 94


RESULT   8
AI767719/c
LOCUS       AI767719      428 bp    mRNA            EST       20-DEC-1999
DEFINITION  wh39a02.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2383082 3',
            mRNA sequence.
ACCESSION   AI767719
VERSION     AI767719.1  GI:5234228
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 428)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 276  Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .428
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2383082"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT      101 a    104 c     57 g    166 t
ORIGIN

  Query Match             4.1%;  Score 25;  DB 24;  Length 428;
  Best Local Similarity   100.0%;  Pred. No. 0.024;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      75 gtggggaacatgctactggcatcta 99
           |||||||||||||||||||||||||
Db     101 GTGGGGAACATGCTACTGGCATCTA 77


RESULT   9
AI609595/c
LOCUS       AI609595      530 bp    mRNA            EST       14-MAY-1999
DEFINITION  tw28d08.x1 NCI_CGAP_Ov35 Homo sapiens cDNA clone IMAGE:2261007 3',
            mRNA sequence.
ACCESSION   AI609595
VERSION     AI609595.1  GI:4618762
```

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 576    Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 404
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2261007"
                     /clone_lib="NCI_CGAP_Ov35"
                     /tissue_type="tumor, 5 pooled (see description)"
                     /lab_host="DH10B"
                     /note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; This library represents the normalized
                     version of NCI_CGAP_Ov23. Cloned unidirectionally.
                     Primer: Oligo dT.  Average insert size 0.86 kb. Tumor
                     types include: mixed Mullerian tumor, papillary serous,
                     clear cell, spindle cell.  All are primary tumors,
                     metastasis positive. Constructed by Life Technologies."
BASE COUNT      131 a    137 c     63 g    198 t      1 others
ORIGIN

  Query Match             4.1%;  Score 25;  DB 22;  Length 530;
  Best Local Similarity   100.0%;  Pred. No. 0.024;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      75 gtggggaacatgctactggcatcta 99
           |||||||||||||||||||||||||
Db     105 GTGGGGAACATGCTACTGGCATCTA 81


RESULT  10
AW771618/c
LOCUS       AW771618      541 bp    mRNA            EST       04-MAY-2000
DEFINITION  hn59c06.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3032170 3',
            mRNA sequence.
ACCESSION   AW771618
VERSION     AW771618.1  GI:7703676
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M D , Ph D
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 463.
FEATURES             Location/Qualifiers
     source          1. .541
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3032170"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT      136 a    136 c     65 g    204 t
ORIGIN

  Query Match             4.1%;  Score 25;  DB 120;  Length 541;
  Best Local Similarity   100.0%;  Pred. No. 0.024;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      75 gtggggaacatgctactggcatcta 99
           |||||||||||||||||||||||||
Db     101 GTGGGGAACATGCTACTGGCATCTA 77


RESULT  11
BE144594
LOCUS       BE144594      323 bp    mRNA            EST       21-JUN-2000
DEFINITION  MR0-HT0167-141199-002-c02 HT0167 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE144594
VERSION     BE144594.1  GI:8607227
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 323)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=MR0-HT0167-141
            199-002-c02&t3=1999-11-14&t4=1)
            Seq primer: puc 18 forward
```

```
          High quality sequence start: 13
          High quality sequence stop: 323.
FEATURES             Location/Qualifiers
     source          1. .323
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0167"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      103 a     75 c     79 g     65 t      1 others
ORIGIN

  Query Match             3.8%;  Score 23;  DB 164;  Length 323;
  Best Local Similarity   100.0%;  Pred. No. 0.29;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    163 agaattatctagccccaaatgtc 185
          |||||||||||||||||||||||
Db    135 AGAATTATCTAGCCCCAAATGTC 157


RESULT  12
B03713
LOCUS       B03713        390 bp    DNA              GSS       13-JUL-1996
DEFINITION  cSRL-185e7-u cSRL flow sorted Chromosome 11 specific cosmid Homo
            sapiens genomic clone cSRL-185e7, DNA sequence.

ACCESSION   B03713
VERSION     B03713.1  GI:1412991
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1  (bases 1 to 390)
  AUTHORS   Evans,G.A., Burbee,D., Davies,C., Hahner,L., Oliver,T., Gilbert,M.,
            Jones,D., Ward,T., Gillilan,E., Schagemann,J., Probst,S. ., Harris
            ,J., DeFord,J., McFarland,J., Burzinski,K., Khan,M., Kupfer,K. and
            Garner,H.R.
  TITLE     Genomic Sequence Sampled Map of Chromosome 11
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Evans GA, Shane Probst
            McDermott Center for Human Growth and Development
            University of Texas Southwestern Medical Center At Dallas
            5323 Harry Hines Blvd, Dallas TX 75235-8591
            Tel: 214-648-1600
            Fax: 214-648-1666
            Email: gevans@utsw.swmed.edu, shane@mcdermott.swmed.edu
            PCR Primers
            FORWARD: ACAAAGAATTATCTAGCCCC
            BACKWARD: AACCACAAACTCAACAACAC
            Seq primer: T7
            Class: cosmid ends
            High quality sequence stop: 390.
FEATURES             Location/Qualifiers
     source          1. .390
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="cSRL-185e7"
                     /clone_lib="cSRL flow sorted Chromosome 11 specific
                     cosmid"
                     /sex="female"
                     /cell_type="chimeric hamster somatic cell hybrid"
                     /note="Vector: sCos-1; Human Chromosome 11 specific cosmid
                     library prepared from flow sorted human  Chromosome 11
                     derived from Chinese Hampster Ovary (CHO) monochromosomal
                     somatic cell hybrid, J1"
BASE COUNT      85 a    111 c     70 g    113 t     11 others
ORIGIN

  Query Match             3.8%;  Score 23;  DB 256;  Length 390;
  Best Local Similarity   100.0%;  Pred. No. 0.29;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    163 agaattatctagccccaaatgtc 185
          |||||||||||||||||||||||
Db     51 AGAATTATCTAGCCCCAAATGTC 73


RESULT  13
AQ282338/c
LOCUS       AQ282338        528 bp    DNA              GSS       27-APR-1999
DEFINITION  RPCI11-79O11.TV RPCI-11 Homo sapiens genomic clone RPCI-11-79O11,
            DNA sequence.

ACCESSION   AQ282338
VERSION     AQ282338.1  GI:3908887
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1  (bases 1 to 528)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Homo sapiens"
                     /db_xref="GDB:7530298"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-79O11"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      162 a     96 c    135 g    135 t
ORIGIN

  Query Match             3.8%;  Score 23;  DB 226;  Length 528;
  Best Local Similarity   100.0%;  Pred. No. 0.29;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    163 agaattatctagccccaaatgtc 185
          |||||||||||||||||||||||
Db    500 AGAATTATCTAGCCCCAAATGTC 478


RESULT  14
R77087
LOCUS       R77087        422 bp    mRNA             EST       06-JUN-1995
DEFINITION  yi64g04.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
```

```
            IMAGE:144054 5' similar to contains LTR4 repetitive element ;, mRNA
            sequence.
ACCESSION   R77087
VERSION     R77087.1  GI:851719
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 422)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 587
            High quality sequence stops: 333
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Putative full length read
            Insert Length: 587   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 333.
FEATURES             Location/Qualifiers
     source          1. .422
                     /organism="Homo sapiens"
                     /db_xref="GDB:553380"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:144054"
                     /clone_lib="Soares placenta Nb2HP"
                     /sex="Female"
                     /dev_stage="placenta obtained at birth (full term)"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo.  "
BASE COUNT      102 a     93 c     88 g    134 t      5 others
ORIGIN

  Query Match             3.5%;  Score 21;  DB 188;  Length 422;
  Best Local Similarity   100.0%;  Pred. No. 3.5;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     245 tgcctttgttctggcttctgt 265
           |||||||||||||||||||||
Db     350 TGCCTTTGTTCTGGCTTCTGT 370


RESULT  15
AL513616/c
LOCUS       AL513616     1018 bp    mRNA            EST       13-FEB-2001
DEFINITION  AL513616 LTI_NFL006_PL2 Homo sapiens cDNA clone XCL0BB001ZE07 5
            prime, mRNA sequence.
ACCESSION   AL513616
VERSION     AL513616.1  GI:12777110
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1018)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .1018
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="XCL0BB001ZE07"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      212 a     308 c     318 g     176 t      4 others
ORIGIN

  Query Match             3.5%;  Score 21;  DB 105;  Length 1018;
  Best Local Similarity   100.0%;  Pred. No. 3.7;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     459 gccctcaggcctgggctgctg 479
           |||||||||||||||||||||
Db      31 GCCCTCAGGCCTGGGCTGCTG 11


Search completed: November  5, 2001, 23:19:10
Job time: 2595 sec
```

GenCore version 4.5

M nucleic - nucleic search, using sw model

un on: November 5, 2001, 22:16:10 ; Search time 1182.83 Seconds
(without alignments)
7924.601 Million cell updates/sec

itle: US-09-656-668-198
erfect score: 606
equence: 1 tgagtttgccccttacccc.....aagcctgttcttgtctgcac 606

coring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

earched: 1344157 seqs, 7733874588 residues

otal number of hits satisfying chosen parameters: 2688314

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	606	100.0	606	10	AX093380	AX093380 Sequence
c	2	104.6	17.3	144014	66	AC021763	AC021763 Homo sapi
	3	104.6	17.3	159150	83	AP002438	AP002438 Homo sapi
	4	104.6	17.3	187200	82	AP001379	AP001379 Homo sapi
c	5	97.2	16.0	152081	86	AC007381	AC007381 Homo sapi
c	6	95.6	15.8	93714	79	AL161661	AL161661 Homo sapi
	7	95.6	15.8	147815	92	HS998H6	AL031687 Human DNA
	8	93.6	15.4	185143	69	AC025861	AC025861 Homo sapi

9	93.6	15.4	208763	79	AL353791	AL353791 Homo sapi
10	91.8	15.1	147929	87	AC016951	AC016951 Homo sapi
11	91	15.0	194313	79	AL353729	AL353729 Homo sapi
12	91	15.0	207746	64	AC017108	AC017108 Homo sapi
13	90.8	15.0	162082	66	AC021051	AC021051 Homo sapi
14	90.8	15.0	180861	74	AC073244	AC073244 Homo sapi
15	90.4	14.9	63020	67	AC022612	AC022612 Homo sapi
16	90.4	14.9	195141	73	AC068045	AC068045 Homo sapi
17	89.6	14.8	163511	90	AL442203	AL442203 Human DNA
18	89.6	14.8	177967	71	AC027806	AC027806 Homo sapi
19	89.6	14.8	186133	79	AL157932	AL157932 Homo sapi
20	89.6	14.8	191831	66	AC021749	AC021749 Homo sapi
21	89.6	14.8	325013	64	AC016817	AC016817 Homo sapi
22	89.2	14.7	155375	92	HS914P14	AL031117 Human DNA
23	88.6	14.6	160814	81	AL449323	AL449323 Homo sapi
24	88.4	14.6	160421	69	AC025005	AC025005 Homo sapi
25	88.4	14.6	190840	79	AL353766	AL353766 Homo sapi
26	88.2	14.6	142756	86	AC005844	AC005844 Homo sapi
27	88.2	14.6	157416	79	AL354914	AL354914 Homo sapi
28	88.2	14.6	182810	67	AC022491	AC022491 Homo sapi
29	87.8	14.5	189072	71	AC040965	AC040965 Homo sapi
30	87.6	14.5	169402	71	AC027715	AC027715 Homo sapi
31	87	14.4	107462	73	AC068208	AC068208 Homo sapi
32	87	14.4	189351	73	AC068588	AC068588 Homo sapi
33	87	14.4	193577	73	AC068025	AC068025 Homo sapi
34	86.8	14.3	110000	80	AL390202_07	Continuation (8 of
35	86.8	14.3	172096	76	AC079864	AC079864 Homo sapi
36	86.8	14.3	218708	66	AC021166	AC021166 Homo sapi
37	86.2	14.2	133202	90	AL139182	AL139182 Human DNA
38	86.2	14.2	153178	61	AC009929	AC009929 Homo sapi
39	86.2	14.2	168293	65	AC018397	AC018397 Homo sapi
40	85.8	14.2	115564	92	HS1170D6	AL030995 Human DNA
41	85.8	14.2	181218	71	AC036222	AC036222 Homo sapi
42	85.8	14.2	184834	88	AC026122	AC026122 Homo sapi
43	85.6	14.1	146652	81	AL445304	AL445304 Homo sapi
44	85.6	14.1	151357	62	AC011573	AC011573 Homo sapi
45	85.6	14.1	157912	89	AL133387	AL133387 Human DNA

ALIGNMENTS

```
ULT   1
93380
US       AX093380      606 bp    DNA                PAT       30-MAR-2001
INITION  Sequence 198 from Patent WO0118046.
ESSION   AX093380
SION     AX093380.1  GI:13509828
WORDS    .
RCE      human.
RGANISM  Homo sapiens
         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
         Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

ERENCE   1  (bases 1 to 606)
UTHORS   Xu,J. and Stolk,J.A.
ITLE     Ovarian tumor sequences and methods of use therefor
OURNAL   Patent: WO 0118046-A 198 15-MAR-2001;
         CORIXA CORPORATION (US)

TURES             Location/Qualifiers
  source          1. .606
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"

E COUNT      175 a    147 c    123 g    161 t
GIN

uery Match           100.0%;  Score 606;  DB 10;  Length 606;
est Local Similarity  100.0%;  Pred. No. 3.3e-168;
atches  606;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

     1 tgagtttgccccccttaccccccatcccagtgaatatttgcaattcctaaagacgtgttttg 60
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1 TGAGTTTGCCCCCTTACCCCCATCCCAGTGAATATTTGCAATTCCTAAAGACGTGTTTTG 60

Qy      61 attgtcacacctgggtggggaacatgctactggcatctaatgcatagagggcagtaatgc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATTGTCACACCTGGGTGGGGAACATGCTACTGGCATCTAATGCATAGAGGGCAGTAATGC 120

Qy     121 tgctaaacatctttcaacgcacaggacagagccccacaaaagagaattatctagccccaa 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGCTAAACATCTTTCAACGCACAGGACAGAGCCCCACAAAAGAGAATTATCTAGCCCCAA 180

Qy     181 atgtccataacactgctgttgagaaaacctaccgcaggatcttactgggcttcataggta 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATGTCCATAACACTGCTGTTGAGAAAACCTACCGCAGGATCTTACTGGGCTTCATAGGTA 240

Qy     241 agcttgcctttgttctggcttctgtagatatataaaataaagacactgcccagtccctcc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AGCTTGCCTTTGTTCTGGCTTCTGTAGATATATAAAATAAAGACACTGCCCAGTCCCTCC 300

Qy     301 ctcaacgtcccgagccagggctcaaggcaattccaataacagtagaatgaacactaaata 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CTCAACGTCCCGAGCCAGGGCTCAAGGCAATTCCAATAACAGTAGAATGAACACTAAATA 360

Qy     361 ttgatttcaaaatctcagcaactagaagaatgaccaaccatcctggttggcctgggactg 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TTGATTTCAAAATCTCAGCAACTAGAAGAATGACCAACCATCCTGGTTGGCCTGGGACTG 420

Qy     421 tcctagttttagcattgaaagtttcaggttccaggaaagccctcaggcctgggctgctgg 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TCCTAGTTTTAGCATTGAAAGTTTCAGGTTCCAGGAAAGCCCTCAGGCCTGGGCTGCTGG 480

Qy     481 tcaccctagcagctgagggactcttcaatacagaattagtctttgtgcactggagatgaa 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TCACCCTAGCAGCTGAGGGACTCTTCAATACAGAATTAGTCTTTGTGCACTGGAGATGAA 540

Qy     541 tatactttaatttgtaacatgtgaaaacatctataaacatctactgaagcctgttcttgt 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TATACTTTAATTTGTAACATGTGAAAACATCTATAAACATCTACTGAAGCCTGTTCTTGT 600

Qy     601 ctgcac 606
           ||||||
Db     601 CTGCAC 606
```

```
RESULT   2
AC021763/c
LOCUS        AC021763   144014 bp    DNA                HTG       13-OCT-2000
DEFINITION   Homo sapiens chromosome 18 clone RP11-56O21 map 18, WORKING DRAFT
             SEQUENCE, 17 unordered pieces.

ACCESSION    AC021763
VERSION      AC021763.3  GI:10800280
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE    1  (bases 1 to 144014)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE      Homo sapiens chromosome 18, clone RP11-56O21
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 144014)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
             Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
             Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
             Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
             DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
             Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
             Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
             Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
             Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
             Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
             McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
             Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
             Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
             Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
```

```
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 13, 2000 this sequence version replaced gi:9117173.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ---------- Project Information
                Center project name: L2274
                Center clone name: 56_O_21
            ---------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 133958 bases at least Q40
                Consensus quality: 139187 bases at least Q30
                Consensus quality: 141302 bases at least Q20
                Insert size: 162000; agarose-fp
                Insert size: 142414; sum-of-contigs
                Quality coverage: 4.3 in Q20 bases; agarose-fp
                Quality coverage: 4.9 in Q20 bases; sum-of-contigs
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7694: contig of 7694 bp in length
            *     7695 7794: gap of      100 bp
            *     7795     8828: contig of 1034 bp in length
            *     8829 8928: gap of      100 bp
            *     8929    11847: contig of 2919 bp in length
            *    11848 11947: gap of      100 bp
            *    11948    13998: contig of 2051 bp in length
            *    13999 14098: gap of      100 bp
            *    14099    18105: contig of 4007 bp in length
            *    18106 18205: gap of      100 bp
            *    18206    22069: contig of 3864 bp in length
            *    22070 22169: gap of      100 bp
            *    22170    25245: contig of 3076 bp in length
            *    25246 25345: gap of      100 bp
            *    25346    28623: contig of 3278 bp in length
            *    28624 28723: gap of      100 bp
            *    28724    33907: contig of 5184 bp in length
            *    33908 34007: gap of      100 bp
            *    34008    39631: contig of 5624 bp in length
            *    39632 39731: gap of      100 bp
            *    39732    45642: contig of 5911 bp in length
            *    45643 45742: gap of      100 bp
            *    45743    53227: contig of 7485 bp in length
            *    53228 53327: gap of      100 bp
            *    53328    63426: contig of 10099 bp in length
            *    63427 63526: gap of      100 bp
            *    63527    75451: contig of 11925 bp in length
            *    75452 75551: gap of      100 bp
            *    75552    92260: contig of 16709 bp in length
            *    92261 92360: gap of      100 bp
            *    92361   114646: contig of 22286 bp in length
            *   114647 114746: gap of      100 bp
            *   114747   144014: contig of 29268 bp in length.
FEATURES             Location/Qualifiers
     source          1   144014
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-56O21"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .7694
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    7795. .8828
                     /note="assembly_fragment"
     misc_feature    8929. .11847
                     /note="assembly_fragment"
     misc_feature    11948. .13998
                     /note="assembly_fragment"
     misc_feature    14099. .18105
                     /note="assembly_fragment"
     misc_feature    18206. .22069
                     /note="assembly_fragment"
     misc_feature    22170. .25245
                     /note="assembly_fragment"
     misc_feature    25346. .28623
                     /note="assembly_fragment"
     misc_feature    28724. .33907
                     /note="assembly_fragment"
     misc_feature    34008. .39631
                     /note="assembly_fragment"
     misc_feature    39732. .45642
                     /note="assembly_fragment"
     misc_feature    45743. .53227
                     /note="assembly_fragment"
     misc_feature    53328. .63426
                     /note="assembly_fragment"
     misc_feature    63527. .75451
                     /note="assembly_fragment"
     misc_feature    75552. .92260
                     /note="assembly_fragment"
     misc_feature    92361. .114646
                     /note="assembly_fragment"
     misc_feature    114747. .144014
                     /note="assembly_fragment"
BASE COUNT    40525 a  31170 c  30024 g  40693 t   1602 others
ORIGIN

  Query Match             17.3%;  Score 104.6;  DB 66;  Length 144014;
  Best Local Similarity   64.9%;  Pred. No. 6.6e-20;
  Matches  155;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy      24 cccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctgggtggggaac 83
           |||||| |||||||| | | |   ||     ||||| |||||||| | ||    |||||
Db 100021 CCCAGTAAATATTTGGCAATGCCCGAGGACATTTTTGGTTGTCACAACGGGAACAGGAAC 99962

Qy      84 atgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacgcaca 143
           ||||||||||||||||||  ||||||  |   |||||||| |||||| | ||  |||||
Db  99961 CTGCTACTGGCATCTAATGAGTAGAGGTGAAGGATGCTGCTGAACATCCTACAGTGCACA 99902

Qy     144 ggacagagccccacaaaagagaattatctagccccaaatgtccataacactgctgttgag 203
           || |||| ||||| || | ||||| ||| ||||| |||| ||  ||   | |  | ||||
Db  99901 GGTCAGAACCCCATAACAAAGAATGATCCAGCCCAAAATATCACTAGTGCCGAGGCTGAG 99842

Qy     204 aaaacctaccgcaggatcttactgggcttcataggtaagcttgcctttgttctggcttc 262
           ||| |    |  |||   || |   |  |||  |   | |    |     |||| |||
Db  99841 AAACCTGTTCTAAGGGCATTGCAAAGTACCATGAGAGTGATGCTGTGAAAGCTGGTTTC 99783


RESULT  3
AP002438
LOCUS       AP002438   159150 bp    DNA              HTG       06-JUN-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-679N11 map 18q21, WORKING
            DRAFT SEQUENCE, 33 unordered pieces.
ACCESSION   AP002438
```

```
VERSION     AP002438.1  GI:8307742
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-679N11.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159150)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 159,150 genomic DNA of 18q21
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 159150)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     ---------- Genome Center
            Center: RIKEN Genomic Sciences Center(GSC)
            Center code: RIKEN
            Web site: http://hgp.gsc.riken.go.jp/
            Contact: hattori@gsc.riken.go.jp
            ---------- Project Information
            Center project name: HumDraft18
            Center clone name: RP11-679N11
            ---------- Summary Statistics
            Sequencing vector: PCR products; 100% of reads
            Chemistry: Dye-terminator ET-amersham; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 141170 bases at least Q40
            Consensus quality: 150107 bases at least Q30
            Consensus quality: 153808 bases at least Q20
            Insert size: 155950; sum-of-contigs
            Quality coverage: 4.52x in Q20 bases; sum-of-contigs
            ----------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            33 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                 1    18093 contig of  18093 bp in length
             18194    34028 contig of  15835 bp in length
             34129    47853 contig of  13725 bp in length
             47954    61375 contig of  13422 bp in length
             61476    69763 contig of   8288 bp in length
             69864    77424 contig of   7561 bp in length
             77525    83516 contig of   5992 bp in length
             83617    89288 contig of   5672 bp in length
             89389    94381 contig of   4993 bp in length
             94482   100220 contig of   5739 bp in length
            100321   105057 contig of   4737 bp in length
            105158   110565 contig of   5408 bp in length
            110666   115962 contig of   5297 bp in length
            116063   119619 contig of   3557 bp in length
            119720   124300 contig of   4581 bp in length
            124401   127169 contig of   2769 bp in length
            127270   129941 contig of   2672 bp in length
            130042   132701 contig of   2660 bp in length
            132802   135223 contig of   2422 bp in length
            135324   138940 contig of   3617 bp in length
            139041   140352 contig of   1312 bp in length
            140453   142867 contig of   2415 bp in length
            142968   144444 contig of   1477 bp in length
            144545   146121 contig of   1577 bp in length
            146222   147660 contig of   1439 bp in length
            147761   149057 contig of   1297 bp in length
            149158   150648 contig of   1491 bp in length
            150749   152399 contig of   1651 bp in length
            152500   154103 contig of   1604 bp in length
            154204   155806 contig of   1603 bp in length
            155907   157204 contig of   1298 bp in length
            157305   158735 contig of   1431 bp in length
            158836   159150 contig of    315 bp in length.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1    18093: contig of 18093 bp in length
*    18094 18193: gap of      100 bp
*    18194    34028: contig of 15835 bp in length
*    34029 34128: gap of      100 bp
*    34129    47853: contig of 13725 bp in length
*    47854 47953: gap of      100 bp
*    47954    61375: contig of 13422 bp in length
*    61376 61475: gap of      100 bp
*    61476    69763: contig of 8288 bp in length
*    69764 69863: gap of      100 bp
*    69864    77424: contig of 7561 bp in length
*    77425 77524: gap of      100 bp
*    77525    83516: contig of 5992 bp in length
*    83517 83616: gap of      100 bp
*    83617    89288: contig of 5672 bp in length
*    89289 89388: gap of      100 bp
*    89389    94381: contig of 4993 bp in length
*    94382 94481: gap of      100 bp
*    94482   100220: contig of 5739 bp in length
*   100221 100320: gap of      100 bp
*   100321   105057: contig of 4737 bp in length
*   105058 105157: gap of      100 bp
*   105158   110565: contig of 5408 bp in length
*   110566 110665: gap of      100 bp
*   110666   115962: contig of 5297 bp in length
*   115963 116062: gap of      100 bp
*   116063   119619: contig of 3557 bp in length
*   119620 119719: gap of      100 bp
*   119720   124300: contig of 4581 bp in length
*   124301 124400: gap of      100 bp
*   124401   127169: contig of 2769 bp in length
*   127170 127269: gap of      100 bp
*   127270   129941: contig of 2672 bp in length
*   129942 130041: gap of      100 bp
*   130042   132701: contig of 2660 bp in length
*   132702 132801: gap of      100 bp
*   132802   135223: contig of 2422 bp in length
*   135224 135323: gap of      100 bp
*   135324   138940: contig of 3617 bp in length
*   138941 139040: gap of      100 bp
*   139041   140352: contig of 1312 bp in length
*   140353 140452: gap of      100 bp
*   140453   142867: contig of 2415 bp in length
*   142868 142967: gap of      100 bp
*   142968   144444: contig of 1477 bp in length
*   144445 144544: gap of      100 bp
*   144545   146121: contig of 1577 bp in length
*   146122 146221: gap of      100 bp
*   146222   147660: contig of 1439 bp in length
*   147661 147760: gap of      100 bp
*   147761   149057: contig of 1297 bp in length
*   149058 149157: gap of      100 bp
*   149158   150648: contig of 1491 bp in length
*   150649 150748: gap of      100 bp
*   150749   152399: contig of 1651 bp in length
*   152400 152499: gap of      100 bp
*   152500   154103: contig of 1604 bp in length
*   154104 154203: gap of      100 bp
*   154204   155806: contig of 1603 bp in length
```

```
          *   155807 155906: gap of      100 bp
          *   155907   157204: contig of 1298 bp in length
          *   157205 157304: gap of      100 bp
          *   157305   158735: contig of 1431 bp in length
          *   158736 158835: gap of      100 bp
          *   158836   159150: contig of 315 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159150
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18q21"
                     /clone="RP11-679N11"
     misc_feature    1. .18093
                     /note="assembly_fragment"
     misc_feature    18194. .34028
                     /note="assembly_fragment"
     misc_feature    34129. .47853
                     /note="assembly_fragment"
     misc_feature    47954. .61375
                     /note="assembly_fragment"
     misc_feature    61476. .69763
                     /note="assembly_fragment"
     misc_feature    69864. .77424
                     /note="assembly_fragment"
     misc_feature    77525. .83516
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    83617. .89288
                     /note="assembly_fragment"
     misc_feature    89389. .94381
                     /note="assembly_fragment"
     misc_feature    94482. .100220
                     /note="assembly_fragment"
     misc_feature    100321. .105057
                     /note="assembly_fragment"
     misc_feature    105158. .110565
                     /note="assembly_fragment"
     misc_feature    110666. .115962
                     /note="assembly_fragment"
     misc_feature    116063. .119619
                     /note="assembly_fragment"
     misc_feature    119720. .124300
                     /note="assembly_fragment"
     misc_feature    124401. .127169
                     /note="assembly_fragment"
     misc_feature    127270. .129941
                     /note="assembly_fragment"
     misc_feature    130042. .132701
                     /note="assembly_fragment"
     misc_feature    132802. .135223

  Query Match             17.3%;  Score 104.6;  DB 83;  Length 159150;
  Best Local Similarity   64.9%;  Pred. No. 6.7e-20;
  Matches  155;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy      24 cccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctgggtggggaac 83
           |||||| ||||||||  | | |   ||     ||||| |||||||| | ||    |||||
Db   24951 CCCAGTAAATATTTGGCAATGCCCGAGGACATTTTTGGTTGTCACAACGGGAACAGGAAC 25010

Qy      84 atgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacgcaca 143
           ||||||||||||||||||  ||||||  |   ||||||| |||||| | ||  |||||
Db   25011 CTGCTACTGGCATCTAATGAGTAGAGGTGAAGGATGCTGCTGAACATCCTACAGTGCACA 25070

Qy     144 ggacagagccccacaaaagagaattatctagccccaaatgtccataacactgctgttgag 203
           || |||| ||||| || | ||||| ||| ||||| |||| ||  ||   | |  | ||||
Db   25071 GGTCAGAACCCCATAACAAAGAATGATCCAGCCCAAAATATCACTAGTGCCGAGGCTGAG 25130

Qy     204 aaaacctaccgcaggatcttactgggcttcataggtaagcttgcctttgttctggcttc 262
           ||| |    |  |||   || |   |   |||  |   | |   |     |||| |||
Db   25131 AAACCTGTTCTAAGGGCATTGCAAAGTACCATGAGAGTGATGCTGTGAAAGCTGGTTTC 25189

RESULT 4
```

```
AP001379
LOCUS       AP001379   187200 bp    DNA              HTG       15-JUL-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-850H3 map 18q12, WORKING
            DRAFT SEQUENCE, 21 unordered pieces.
ACCESSION   AP001379
VERSION     AP001379.3  GI:9229958
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-850H3.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 187200)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 187,200 genomic DNA of 18q12
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 187200)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT     On Jul 15, 2000 this sequence version replaced gi:8117292.
            ---------- Genome Center
               Center: RIKEN Genomic Sciences Center(GSC)
               Center code: RIKEN
               Web site: http://hgp.gsc.riken.go.jp/
               Contact: hattori@gsc.riken.go.jp
            ---------- Project Information
               Center project name: HumDraft18
               Center clone name: RP11-850H3
            ---------- Summary Statistics
               Sequencing vector: PCR products; 100% of reads
               Chemistry: Dye-terminator ET-amersham; 100% of reads
               Assembly program: Phrap; version 0.990329
               Consensus quality: 179971 bases at least Q40
               Consensus quality: 182832 bases at least Q30
               Consensus quality: 184194 bases at least Q20
               Insert size: 185200; sum-of-contigs
               Quality coverage: 9.46x in Q20 bases; sum-of-contigs
            ----------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            21 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1    33581 contig of  33581 bp in length
              33682    58093 contig of  24412 bp in length
              58194    77625 contig of  19432 bp in length
              77726    97031 contig of  19306 bp in length
              97132   111990 contig of  14859 bp in length
             112091   125465 contig of  13375 bp in length
             125566   135454 contig of   9889 bp in length
             135555   144628 contig of   9074 bp in length
             144729   151062 contig of   6334 bp in length
             151163   155511 contig of   4349 bp in length
             155612   160186 contig of   4575 bp in length
             160287   164088 contig of   3802 bp in length
             164189   167739 contig of   3551 bp in length
             167840   171188 contig of   3349 bp in length
             171289   174981 contig of   3693 bp in length
             175082   177828 contig of   2747 bp in length
             177929   180660 contig of   2732 bp in length
             180761   183192 contig of   2432 bp in length
             183293   185028 contig of   1736 bp in length
             185129   185875 contig of    747 bp in length
             185976   187200 contig of   1225 bp in length.
          * NOTE: This is a 'working draft' sequence. It currently
```

```
        * consists of 21 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *        1    33581: contig of 33581 bp in length
        *    33582 33681: gap of      100 bp
        *    33682    58093: contig of 24412 bp in length
        *    58094 58193: gap of      100 bp
        *    58194    77625: contig of 19432 bp in length
        *    77626 77725: gap of      100 bp
        *    77726    97031: contig of 19306 bp in length
        *    97032 97131: gap of      100 bp
        *    97132   111990: contig of 14859 bp in length
        *   111991 112090: gap of      100 bp
        *   112091   125465: contig of 13375 bp in length
        *   125466 125565: gap of      100 bp
        *   125566   135454: contig of 9889 bp in length
        *   135455 135554: gap of      100 bp
        *   135555   144628: contig of 9074 bp in length
        *   144629 144728: gap of      100 bp
        *   144729   151062: contig of 6334 bp in length
        *   151063 151162: gap of      100 bp
        *   151163   155511: contig of 4349 bp in length
        *   155512 155611: gap of      100 bp
        *   155612   160186: contig of 4575 bp in length
        *   160187 160286: gap of      100 bp
        *   160287   164088: contig of 3802 bp in length
        *   164089 164188: gap of      100 bp
        *   164189   167739: contig of 3551 bp in length
        *   167740 167839: gap of      100 bp
        *   167840   171188: contig of 3349 bp in length
        *   171189 171288: gap of      100 bp
        *   171289   174981: contig of 3693 bp in length
        *   174982 175081: gap of      100 bp
        *   175082   177828: contig of 2747 bp in length
        *   177829 177928: gap of      100 bp
        *   177929   180660: contig of 2732 bp in length
        *   180661 180760: gap of      100 bp
        *   180761   183192: contig of 2432 bp in length
        *   183193 183292: gap of      100 bp
        *   183293   185028: contig of 1736 bp in length
        *   185029 185128: gap of      100 bp
        *   185129   185875: contig of 747 bp in length
        *   185876 185975: gap of      100 bp
        *   185976   187200: contig of 1225 bp in length.
FEATURES             Location/Qualifiers
     source          1. .187200
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18q12"
                     /clone="RP11-850H3"
     misc_feature    1. .33581
                     /note="assembly_fragment"
     misc_feature    33682. .58093
                     /note="assembly_fragment"
     misc_feature    58194. .77625
                     /note="assembly_fragment"
     misc_feature    77726. .97031
                     /note="assembly_fragment clone_end:SP6 vector_side:right"
     misc_feature    97132. .111990
                     /note="assembly_fragment"
     misc_feature    112091. .125465
                     /note="assembly_fragment"
     misc_feature    125566. .135454
                     /note="assembly_fragment"
     misc_feature    135555. .144628
                     /note="assembly_fragment"
     misc_feature    144729. .151062
                     /note="assembly_fragment"
     misc_feature    151163. .155511
                     /note="assembly_fragment"
     misc_feature    155612. .160186
                     /note="assembly_fragment"
     misc_feature    160287. .164088
                     /note="assembly_fragment"
     misc_feature    164189. .167739
                     /note="assembly_fragment"
     misc_feature    167840. .171188
                     /note="assembly_fragment"
     misc_feature    171289. .174981
                     /note="assembly_fragment"
     misc_feature    175082. .177828
                     /note="assembly_fragment"
     misc_feature    177929. .180660
                     /note="assembly_fragment"
     misc_feature    180761. .183192
                     /note="assembly_fragment"
     misc_feature    183293. .185028
                     /note="assembly_fragment"
     misc_feature    185129. .185875
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    185976. .187200
                     /note="assembly_fragment"
BASE COUNT    51230 a  39996 c  41403 g  52571 t   2000 others
ORIGIN

  Query Match           17.3%;  Score 104.6;  DB 82;  Length 187200;
  Best Local Similarity  64.9%;  Pred. No. 6.8e-20;
  Matches  155;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy      24 cccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctgggtggggaac 83
           ||||||  ||||||||  | | |   ||     ||||| |||||||| | ||    |||||
Db 116028 CCCAGTAAATATTTGGCAATGCCCGAGGACATTTTTGGTTGTCACAACGGGAACAGGAAC 116087

Qy      84 atgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacgcaca 143
           ||||||||||||||||||  ||||||  |   ||||||||  ||||||  | ||  |||||
Db 116088 CTGCTACTGGCATCTAATGAGTAGAGGTGAAGGATGCTGCTGAACATCCTACAGTGCACA 116147

Qy     144 ggacagagccccacaaaagagaattatctagccccaaatgtccataacactgctgttgag 203
           || |||| ||||| || | ||||| ||| ||||| |||| ||  ||   | |  | ||||
Db 116148 GGTCAGAACCCCATAACAAAGAATGATCCAGCCCAAAATATCACTAGTGCCGAGGCTGAG 116207

Qy     204 aaaacctaccgcaggatcttactgggcttcataggtaagcttgcctttgttctggcttc 262
           ||| |    |  |||   || |   |   |||  |   | |    |     |||| |||
Db 116208 AAACCTGTTCTAAGGGCATTGCAAAGTACCATGAGAGTGATGCTGTGAAAGCTGGTTTC 116266


RESULT   5
AC007381/c
LOCUS       AC007381   152081 bp    DNA             PRI       30-SEP-2000
DEFINITION  Homo sapiens BAC clone RP11-158I21 from 2, complete sequence.
ACCESSION   AC007381
VERSION     AC007381.3  GI:6604532
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 152081)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 152081)
  AUTHORS   Kozlowicz,A., Wohldmann,P. and Harper,M.
  TITLE     The sequence of Homo sapiens BAC clone RP11-158I21
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 152081)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-1999) Genome Sequencing Center, Washington
```

```
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 152081)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 152081)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 152081)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Dec 20, 1999 this sequence version replaced gi:5001522.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0158I21
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is RP11-119H15.  Actual start of
            this clone is at base position 1 of RP11-158I21.

            The sequence RP11-158I21 contains a dinucleotide (CT) repeat from
            base positions 6270 to 7430 in which the exact length is unknown.
            A PCR product from clone DNA confirms the approximate size of the
            region with band sizes of 2527 real, 2508 insilico for EcoRV; and
            6293 real, 6234 insilico for HindIII.

FEATURES            Location/Qualifiers
     source         1. .152081
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="2"
                    /map="2"
                    /clone="RP11-158I21"
                    /clone_lib="RPCI-11"
     misc_feature   97. .567
                    /note="match to EST N54976 (NID:g1196296) yv38h03.s1"
     repeat_region  1252. .1426
                    /rpt_family="(CA)n"
     repeat_region  2507. .2687
                    /rpt_family="MER1_type"
     repeat_region  3468. .3596
                    /rpt_family="MIR"
     repeat_region  3631. .3662
                    /rpt_family="(CAAAAA)n"
     misc_feature   5619. .5723
                    /note="similar to Mus musculus EST AI255229 (NID:g3862754)
                    ui32h10.y1"
     misc_feature   5637. .5723
                    /note="similar to Mus musculus EST AA254455 (NID:g1889058)
                    va16e12.r1"
     misc_feature   5641
                    /note="match to EST R88177 (NID:g946990) ym90e05.r1"
     misc_feature   5669. .5723
                    /note="similar to Mus musculus EST AW123279
                    (NID:g6098809)"
     misc_feature   5684. .5723
                    /note="match to EST AI081350 (NID:g3418142) ox76e08.x1"
     misc_feature   5698. .5716
                    /note="match to EST R88177 (NID:g946990) ym90e05.r1"
     repeat_region  6293. .7016
                    /rpt_family="(CCA)n"
     repeat_region  7023. .7383
                    /rpt_family="(CCA)n"
     repeat_region  8687. .8848
                    /rpt_family="MIR"
     repeat_region  8934. .8959
                    /rpt_family="(TAAAA)n"
     repeat_region  11635. .11696
                    /rpt_family="MIR"
     repeat_region  11845. .11891
                    /rpt_family="L2"
     misc_feature   12021. .12411
                    /note="match to EST R88177 (NID:g946990) ym90e05.r1"
     misc_feature   12031. .12173
                    /note="similar to Mus musculus EST AI255229 (NID:g3862754)
                    ui32h10.y1"
     misc_feature   12031. .12173
                    /note="match to EST AI081350 (NID:g3418142) ox76e08.x1"
     misc_feature   12031. .12173
                    /note="similar to Mus musculus EST AW123279
                    (NID:g6098809)"
     misc_feature   12031. .12173
                    /note="similar to Mus musculus EST AA254455 (NID:g1889058)
                    va16e12.r1"
     misc_feature   12073. .12173
                    /note="similar to Mus musculus EST AA183217 (NID:g1766873)
                    mt80h06.r1"
     misc_feature   12091. .12173
                    /note="similar to Mus musculus EST AA002793 (NID:g1446526)
                    mg41g01.r1"
     misc_feature   12101. .12173
                    /note="similar to Mus musculus EST AA726437 (NID:g2744144)
                    vu92b11.r1"
     misc_feature   12117. .12173
                    /note="match to EST AI560419 (NID:g4510760) tn09h07.x1"
     misc_feature   12124. .12173
                    /note="similar to Mus musculus EST AA003295 (NID:g1446732)
                    mg48f11.r1"
     misc_feature   12134. .12173
                    /note="match to EST AA236279 (NID:g1858428) zr54f05.s1"
     misc_feature   12598. .13767
                    /note="CpG_island (%GC=66.2, o/e=0.80, #CpGs=110)"
     misc_feature   12974. .13396
                    /note="match to EST AW157755 (NID:g6229156) au86d06.x1"
     repeat_region  12983. .13055
```

```
                     /rpt_family="GA-rich"
     misc_feature    13467. .13930
                     /note="similar to Danio rerio EST AI958657 (NID:g5751370)
                     fc94g02.y1"
     misc_feature    13470. .13843
                     /note="similar to Mus musculus EST AI060775 (NID:g3336198)
                     ub43e04.r1"
     misc_feature    13728. .14181
                     /note="similar to Danio rerio EST AW116140 (NID:g6082478)
                     fi09f10.x1"
     repeat_region   14169. .14189
                     /rpt_family="(CAAA)n"
     repeat_region   14238. .14291
                     /rpt_family="(T)n"
     misc_feature    14240. .14639
                     /note="similar to Rattus norvegicus EST AA964818
                     (NID:g4279692)"
     misc_feature    14249. .14403
                     /note="similar to Rattus norvegicus EST AI704282
                     (NID:g4992182)"
     misc_feature    14266. .14451
                     /note="match to EST AA504508 (NID:g2240668) aa60g08.s1"
     misc_feature    14287. .14414
                     /note="similar to Danio rerio EST AW116140 (NID:g6082478)
                     fi09f10.x1"
     misc_feature    14405. .14886
                     /note="similar to Rattus sp. EST AW143415 (NID:g6163418)"
     repeat_region   14497. .14526
                     /rpt_family="(GA)n"
     misc_feature    14541. .14909
                     /note="similar to Rattus norvegicus EST AI029866
                     (NID:g4301527)"

  Query Match           16.0%;  Score 97.2;  DB 86;  Length 152081;
  Best Local Similarity  72.2%;  Pred. No. 1e-17;
  Matches  140;  Conservative    0;  Mismatches   53;  Indels    1;  Gaps    1;

Qy      17 cccccatcccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctgggt 76
           | ||  |||||  ||   || |||||  ||| |||| | ||||| |||||||| || ||
Db   70146 CTCCACTCCCAAGGACATTTGGCAATGTCTAGAGACAT-TTTTGGTTGTCACAACTCGGG 70088

Qy      77 ggggaacatgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttca 136
           || |     || |||||||||||| || ||||||| | |  |||||||||||| ||| | |
Db   70087 GGAGGGGGTGATACTGGCATCTAGTGGATAGAGGCCCGGGATGCTGCTAAAAATCCTATA 70028

Qy     137 acgcacaggacagagccccacaaaagagaattatctagccccaaatgtccataacactgc 196
           | |||||||||||| |||||||| | |||||| ||  |||| |||||||| | |   | ||
Db   70027 ATGCACAGGACAGACCCCCACAACAAAGAATTCTCCTGCCCAAAATGTCAACAGTGCAGA 69968

Qy     197 tgttgagaaaacct 210
             ||| ||||   ||
Db   69967 GGTTAAGAACCTCT 69954


RESULT   6
AL161661/c
LOCUS       AL161661    93714 bp    DNA             HTG       20-JAN-2001
DEFINITION  Homo sapiens chromosome 20 clone RP5-984I23, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL161661
VERSION     AL161661.1  GI:7263577
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 93714)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     ---------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ---------- Project Information
            Center project name: dJ984I23
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 93681 bases at least Q40
            Consensus quality: 93706 bases at least Q30
            Consensus quality: 93713 bases at least Q20
            Insert size: 93714; sum-of-contigs
            Insert size: 109355; 1.0% error; agarose-fp
            Quality coverage: 8.25x in Q20 bases; sum-of-contigs Quality
            coverage: 7.07x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .93714
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP5-984I23"
                     /clone_lib="RPCI-5"
     misc_feature    1. .93714
                     /note="assembly_fragment:00029"
BASE COUNT    24148 a  22562 c  22969 g  24035 t
ORIGIN

  Query Match           15.8%;  Score 95.6;  DB 79;  Length 93714;
  Best Local Similarity  75.8%;  Pred. No. 2.9e-17;
  Matches  144;  Conservative    0;  Mismatches   44;  Indels    2;  Gaps    2;

Qy      20 ccatcccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctgggtggg 79
           ||| ||  | ||   || |||||  ||| ||| || || || || |||||||| |||||| ||
Db   44680 CCACCCAGGGGACATTTAGCAATGACTAGAGATGT-TTCTGGTTGTCACAACTGGGT-GG 44623

Qy      80 gaacatgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacg 139
           ||   |||| |||||||||| ||  |||||| ||| ||||||||||||||  | ||| |
Db   44622 GAGGGTGCTGCTGGCATCTAGTGGGTAGAGGCCAGGAATGCTGCTAAACATTCTACAATG 44563

Qy     140 cacaggacagagccccacaaaagagaattatctagccccaaatgtccataacactgctgt 199
           | | ||||||  | |||| | | |||||||||| |||||||||||||| |||   ||| ||
Db   44562 CCCGGGACAGCCCTCCACGACAAAGAATTATCCAGCCCCAAATGTCAATAGAGCTGAGGT 44503

Qy     200 tgagaaaacc 209
           ||||||| ||
Db   44502 TGAGAAACCC 44493


RESULT   7
HS998H6
LOCUS       HS998H6    147815 bp    DNA             PRI       04-APR-2001
DEFINITION  Human DNA sequence from clone RP5-998H6 on chromosome 20q13.1.
            Contains the gene for the ortholog of rat PB-Cadherin, ESTs, STSs,
            GSS, two CpG islands and genomic marker D20S17, complete sequence.
ACCESSION   AL031687
VERSION     AL031687.17  GI:13366276
KEYWORDS    HTG; CpG island; D20S17; PB-Cadherin.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 147815)
  AUTHORS   Laird,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
```

```
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: cloneregquest@sanger.ac.uk
COMMENT     On Mar 16, 2001 this sequence version replaced gi:10198608.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            RP5-998H6 is from the library RPCI-5 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP5-998H6 The true
            right end of clone RP4-599F21 is at 29571 in this sequence. This
            sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest.
FEATURES            Location/Qualifiers
     source         1. .147815
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="20"
                    /map="q13.1"
                    /clone="RP5-998H6"
                    /clone_lib="RPCI-5"
     repeat_region  85. .149
                    /note="MIR repeat: matches 60. .125 of consensus"
     repeat_region  865. .905
                    /note="MER91 repeat: matches 153. .195 of consensus"
     repeat_region  1141. .1358
                    /note="MIR repeat: matches 27. .250 of consensus"
     repeat_region  1959. .2058
                    /note="L2 repeat: matches 2625. .2740 of consensus"
     repeat_region  2081. .2176
                    /note="L2 repeat: matches 2612. .2710 of consensus"
     repeat_region  2303. .2442
                    /note="7 copies 20 mer 81% conserved"
     repeat_region  3585. .3790
                    /note="MIR repeat: matches 20. .239 of consensus"
     repeat_region  4704. .4749
                    /note="23 copies 2 mer gt 84% conserved"
     misc_feature   complement(4990. .5548)
                    /note="match: GSS: Em:AQ602943"
     repeat_region  5125. .5354
                    /note="115 copies 2 mer tg 62% conserved"
     repeat_region  5151. .5390
                    /note="12 copies 20 mer 60% conserved"
     repeat_region  5397. .5442
                    /note="23 copies 2 mer gt 84% conserved"
     repeat_region  5645. .5830
                    /note="MLT1B repeat: matches 1. .110 of consensus"
     repeat_region  5831. .6141
                    /note="AluSq repeat: matches 1. .313 of consensus"
     repeat_region  6142. .6394
                    /note="MLT1B repeat: matches 110. .390 of consensus"
     repeat_region  6855. .7055
                    /note="MIR repeat: matches 59. .260 of consensus"
     repeat_region  7129. .7544
                    /note="L2 repeat: matches 1529. .1989 of consensus"
     repeat_region  7601. .8077
                    /note="MLT1J repeat: matches 34. .514 of consensus"
     repeat_region  8089. .8174
                    /note="L2 repeat: matches 2066. .2155 of consensus"
     repeat_region  8183. .8364
                    /note="MER53 repeat: matches 1. .182 of consensus"
     repeat_region  10532. .10665
                    /note="MIR repeat: matches 80. .220 of consensus"
     repeat_region  10708. .10764
                    /note="MIR repeat: matches 193. .251 of consensus"
     repeat_region  11506. .11687
                    /note="MIR repeat: matches 29. .223 of consensus"
     repeat_region  11702. .11810
                    /note="MIR repeat: matches 82. .190 of consensus"
     repeat_region  11813. .12111
                    /note="AluJo repeat: matches 1. .312 of consensus"
     repeat_region  12113. .12306
                    /note="97 copies 2 mer aa 66% conserved"
     repeat_region  12123. .12302
                    /note="9 copies 20 mer 69% conserved"
     repeat_region  12593. .12701
                    /note="MIR repeat: matches 64. .177 of consensus"
     repeat_region  13003. .13232
                    /note="L2 repeat: matches 2266. .2499 of consensus"
     repeat_region  13372. .13455
                    /note="MIR repeat: matches 72. .161 of consensus"
     repeat_region  13574. .13780
                    /note="MIR repeat: matches 56. .262 of consensus"
     repeat_region  14279. .14479
                    /note="MIR repeat: matches 65. .262 of consensus"
     polyA_site     complement(14281)
                    /gene="dJ998H6.1"
     mRNA           complement(join(14281. .15619,18488. .18739,27130. .27247,
                    27368. .27489,39965. .40101,50849. .51102,53537. .53730,
                    57368. .57535,68050. .68169,81505. .81799,91582. .92237))
                    /gene="dJ998H6.1"
                    /note="match: cDNAs: Em:D83348 Em:D83349 Em:AB019618
                    Em:D82029 Em:D42150 Em:AF007116 Em:D25290 Em:D31784
                    Em:E07383 Em:D21254 Em:E07382 Em:D21255 Em:L34056
                    Em:L33477 Em:L34057 Em:X77557 Em:D31963 Em:D21253
                    Em:E07381 Em:D42149 Em:U59325 Em:X85330 Em:AB000512
                    Em:AF002983 Em:Y13919 Em:AF035300"
                    /evidence=not_experimental
                    /product="dJ998H6.1 (ortholog of rat PB-Cadherin)"
     gene           complement(14281. .92237)
                    /gene="dJ998H6.1"
     polyA_signal   complement(14299. .14304)
                    /gene="dJ998H6.1"
     misc_feature   14968. .15565
                    /note="CpG island"
                    /evidence=not_experimental
     CDS            complement(join(15048. .15619,18488. .18739,27130. .27247,
                    27368. .27489,39965. .40101,50849. .51102,53537. .53730,
                    57368. .57535,68050. .68169,81505. .81799,91582. .91836))
                    /gene="dJ998H6.1"
                    /note="match: proteins: Tr:Q63315 Tr:Q63561 Tr:Q90763
                    Sw:Q13634 Sw:P79995 Tr:Q15066 Tr:Q90762 Sw:P55285
                    Sw:P55280 Sw:P97326 Tr:Q9JIW2 Sw:P55287 Sw:P55288
                    Tr:O93319 Tr:Q9Z0M3 Tr:Q9WTP5 Tr:O43205 Tr:Q9ULB2
                    Sw:P55286 Tr:O54800 Sw:P97291 Sw:P55289 Tr:Q9ULB5
                    Tr:Q9ULB4"
                    /codon_start=1
                    /evidence=not_experimental
                    /product="dJ998H6.1 (ortholog of rat PB-Cadherin)"
                    /protein_id="CAB51587.2"
                    /db_xref="GI:13366277"
                    /db_xref="SPTREMBL:Q9UJ99"
                    /translation="MRPRPEGRGLRAGVALSPALLLLLLPPPPTLLGRLWAAGTPSP
                    SAPGARQDGALGAGRVKRGWVWNQFFVVEEYTGTEPLYVGKIHSDSDEGDGAIKYTIS
                    GEGAGTIFLIDELTGDIHAMERLDREQKTFYTLRAQARDRATNRLLEPESEFIIKVQD
                    INDSEPRFLHGPYIGSVAELSPTGTSVMQVMASDADDPTYGSSARLVYSVLDGEHHFT
                    VDPKTGVIRTAVPDLDRESQERYEVVIQATDMAGQLGGLSGSTTVTIVVTDVNDNPPR
                    FPQKMYQFSIQESAPIGTAVGRVKAEDSDVGENTDMTYHLKDESSSGGDVFKVTTDSD
```

```
                     TQEAIIVVQKRLDFESQPVHTVILEALNKFVDPRFADLGTFRDQAIVRVAVTDVDEPP
                     EFRPPSGLLEVQEDAQVGSLVGVVTARDPDAANRPVRYAIDRESDLDQIFDIDADTGA
                     IVTGKGLDRETAGWHNITVLAMEADNHAQLSRASLRIRILDVNDNPPELATPYEAAVC
                     EDAKPGQLIQTISVVDRDEPQGGHRFYFRLVPEAPSNPHFSLLDIQDNTAAVHTQHVG
                     FNRQEQDVFFLPILVVDSGPPTLSSTGTLTIRICGCDSSGTIQSCNTTAFVMAASLSP
                     GALIALLVCVLILVVLVLLILTLRRHHKSHLSSDEDEDMRDNVIKYNDEGGGEQDTEA
                     YDMSALRSLYDFGELKGGDGGGSAGGGAGGGSGGGAGSPPQAHLPSERHSLPQGPPSP
                     EPDFSVFRDFISRKVALADGDLSVPPYDAFQTYAFEGADSPAASLSSLHSGSSGSEQD
                     FAYLSSWGPRFRPLAALYAGHRGDDEAQAS"
repeat_region        15352. .15451
                     /note="50 copies 2 mer gc 61% conserved"
repeat_region        16084. .16209
                     /note="L2 repeat: matches 2430. .2552 of consensus"
repeat_region        16210. .16371
                     /note="MER5A repeat: matches 4. .187 of consensus"
repeat_region        16372. .16519
                     /note="L2 repeat: matches 2552. .2710 of consensus"
repeat_region        16559. .16989
                     /note="L2 repeat: matches 1421. .1884 of consensus"
repeat_region        18065. .18163
                     /note="MIR repeat: matches 85. .192 of consensus"
repeat_region        20192. .20363
                     /note="MIR repeat: matches 66. .262 of consensus"
repeat_region        20430. .20528
                     /note="L2 repeat: matches 2370. .2473 of consensus"
repeat_region        21074. .21232
                     /note="MIR repeat: matches 91. .255 of consensus"
repeat_region        21282. .21457
                     /note="AluJ repeat: matches 121. .294 of consensus"
repeat_region        21469. .21504
                     /note="18 copies 2 mer tg 100% conserved"
repeat_region        21505. .21616
                     /note="AluJ repeat: matches 2. .120 of consensus"
repeat_region        21618. .22121
                     /note="L2 repeat: matches 2213. .2745 of consensus"
repeat_region        22141. .22340
                     /note="10 copies 20 mer 75% conserved"
repeat_region        22197. .22346
                     /note="75 copies 2 mer tt 66% conserved"

 Query Match              15.8%;  Score 95.6;  DB 92;  Length 147815;
 Best Local Similarity    75.8%;  Pred. No. 3e-17;
 Matches  144;  Conservative    0;  Mismatches   44;  Indels    2;  Gaps    2;

Qy     20 ccatcccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctgggtggg 79
          ||| ||  | ||   || |||||  ||| ||| || || || |||||||| |||||| ||
Db 130835 CCACCCAGGGGACATTTAGCAATGACTAGAGATGT-TTCTGGTTGTCACAACTGGGT-GG 130892

Qy     80 gaacatgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacg 139
          ||   |||| |||||||||| ||  |||||| ||| ||||||||||||||  | ||| |
Db 130893 GAGGGTGCTGCTGGCATCTAGTGGGTAGAGGCCAGGAATGCTGCTAAACATTCTACAATG 130952

Qy    140 cacaggacagagccccacaaaagagaattatctagccccaaatgtccataacactgctgt 199
          | | ||||||  | |||| | | |||||||||| |||||||||||| |||   |||  ||
Db 130953 CCCGGGACAGCCCTCCACGACAAAGAATTATCCAGCCCCAAATGTCAATAGAGCTGAGGT 131012

Qy    200 tgagaaaacc 209
          ||||||| ||
Db 131013 TGAGAAACCC 131022


RESULT   8
AC025861
LOCUS       AC025861   185143 bp    DNA              HTG       26-AUG-2000
DEFINITION  Homo sapiens clone RP11-766D10, WORKING DRAFT SEQUENCE, 44
            unordered pieces.
ACCESSION   AC025861
VERSION     AC025861.3  GI:9929758
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1  (bases 1 to 185143)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-766D10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185143)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 26, 2000 this sequence version replaced gi:9887743.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L7554
                Center clone name: 766_D_10
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 163484 bases at least Q40
                Consensus quality: 174463 bases at least Q30
                Consensus quality: 178842 bases at least Q20
                Insert size: 200000; agarose-fp
                Insert size: 180843; sum-of-contigs
                Quality coverage: 2.9 in Q20 bases; agarose-fp
                Quality coverage: 3.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 44 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1      200: contig of 200 bp in length
            *     201 300: gap of      100 bp
            *     301     1640: contig of 1340 bp in length
            *    1641 1740: gap of      100 bp
            *    1741     3026: contig of 1286 bp in length
            *    3027 3126: gap of      100 bp
            *    3127     4675: contig of 1549 bp in length
            *    4676 4775: gap of      100 bp
            *    4776     5775: contig of 1000 bp in length
            *    5776 5875: gap of      100 bp
            *    5876     7301: contig of 1426 bp in length
            *    7302 7401: gap of      100 bp
```

```
*     7402     8749: contig of 1348 bp in length
*     8750 8849: gap of      100 bp
*     8850    10630: contig of 1781 bp in length
*    10631 10730: gap of      100 bp
*    10731    12287: contig of 1557 bp in length
*    12288 12387: gap of      100 bp
*    12388    14291: contig of 1904 bp in length
*    14292 14391: gap of      100 bp
*    14392    15981: contig of 1590 bp in length
*    15982 16081: gap of      100 bp
*    16082    17373: contig of 1292 bp in length
*    17374 17473: gap of      100 bp
*    17474    19023: contig of 1550 bp in length
*    19024 19123: gap of      100 bp
*    19124    20568: contig of 1445 bp in length
*    20569 20668: gap of      100 bp
*    20669    22533: contig of 1865 bp in length
*    22534 22633: gap of      100 bp
*    22634    24930: contig of 2297 bp in length
*    24931 25030: gap of      100 bp
*    25031    27214: contig of 2184 bp in length
*    27215 27314: gap of      100 bp
*    27315    30058: contig of 2744 bp in length
*    30059 30158: gap of      100 bp
*    30159    31989: contig of 1831 bp in length
*    31990 32089: gap of      100 bp
*    32090    35144: contig of 3055 bp in length
*    35145 35244: gap of      100 bp
*    35245    37733: contig of 2489 bp in length
*    37734 37833: gap of      100 bp
*    37834    40275: contig of 2442 bp in length
*    40276 40375: gap of      100 bp
*    40376    42771: contig of 2396 bp in length
*    42772 42871: gap of      100 bp
*    42872    46769: contig of 3898 bp in length
*    46770 46869: gap of      100 bp
*    46870    49648: contig of 2779 bp in length
*    49649 49748: gap of      100 bp
*    49749    53828: contig of 4080 bp in length
*    53829 53928: gap of      100 bp
*    53929    58678: contig of 4750 bp in length
*    58679 58778: gap of      100 bp
*    58779    63068: contig of 4290 bp in length
*    63069 63168: gap of      100 bp
*    63169    67733: contig of 4565 bp in length
*    67734 67833: gap of      100 bp
*    67834    72886: contig of 5053 bp in length
*    72887 72986: gap of      100 bp
*    72987    76848: contig of 3862 bp in length
*    76849 76948: gap of      100 bp
*    76949    82756: contig of 5808 bp in length
*    82757 82856: gap of      100 bp
*    82857    88725: contig of 5869 bp in length
*    88726 88825: gap of      100 bp
*    88826    95518: contig of 6693 bp in length
*    95519 95618: gap of      100 bp
*    95619   101325: contig of 5707 bp in length
*   101326 101425: gap of      100 bp
*   101426   106203: contig of 4778 bp in length
*   106204 106303: gap of      100 bp
*   106304   111060: contig of 4757 bp in length
*   111061 111160: gap of      100 bp
*   111161   119653: contig of 8493 bp in length
*   119654 119753: gap of      100 bp
*   119754   127901: contig of 8148 bp in length
*   127902 128001: gap of      100 bp
*   128002   136682: contig of 8681 bp in length
*   136683 136782: gap of      100 bp
*   136783   146372: contig of 9590 bp in length
*   146373 146472: gap of      100 bp
*   146473   157184: contig of 10712 bp in length
*   157185 157284: gap of      100 bp
*   157285   170174: contig of 12890 bp in length
*   170175 170274: gap of      100 bp
*   170275   185143: contig of 14869 bp in length.

FEATURES             Location/Qualifiers
     source          1. .185143
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-766D10"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .200
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    301. .1640
                     /note="assembly_fragment"
     misc_feature    1741. .3026
                     /note="assembly_fragment"
     misc_feature    3127. .4675
                     /note="assembly_fragment"
     misc_feature    4776. .5775
                     /note="assembly_fragment"
     misc_feature    5876. .7301
                     /note="assembly_fragment"
     misc_feature    7402. .8749
                     /note="assembly_fragment"
     misc_feature    8850. .10630
                     /note="assembly_fragment"
     misc_feature    10731. .12287
                     /note="assembly_fragment"
     misc_feature    12388. .14291
                     /note="assembly_fragment"
     misc_feature    14392. .15981
                     /note="assembly_fragment"
     misc_feature    16082. .17373
                     /note="assembly_fragment"
     misc_feature    17474. .19023
                     /note="assembly_fragment"
     misc_feature    19124. .20568
                     /note="assembly_fragment"
     misc_feature    20669. .22533
                     /note="assembly_fragment"
     misc_feature    22634. .24930
                     /note="assembly_fragment"
     misc_feature    25031. .27214
                     /note="assembly_fragment"

Query Match          15.4%;  Score 93.6;  DB 69;  Length 185143;
Best Local Similarity  66.2%;  Pred. No. 1.2e-16;
Matches  135;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy      30 gaatatttgcaattcctaaagacgtgttttgattgtcacacctgggtggggaacatgcta 89
           ||   |||||||||  ||  || | | ||  | | |   |   |||  |||   ||||||
Db   35568 GACATTTTGCAATATCTGGAGGCATTTTCAGTTATTATAAGTGGGGCTGGGTGGATGCTG 35627

Qy      90 ctggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacgcacaggacag 149
           ||||||||| ||| | ||||  |||  |||||| |||||||| |  || ||||||||||
Db   35628 CTGGCATCTCATGGAGAGAGACCAGAGATGCTGTTAAACATCCTATAATGCACAGGACAG 35687

Qy     150 agccccacaaaagagaattatctagccccaaatgtccataacactgctgttgagaaaacc 209
           ||||||||| | |||||||||| ||||  |||||||  ||   |||  | |||||||  |
Db   35688 GCCCCCACAACACAGAATTATCCAGCCTAAAATGTCACTAGTGCTGAGGGTGAGAAACTC 35747

Qy     210 taccgcaggatcttactgggcttc 233
           |  | ||     | | ||   |||
Db   35748 TGACACATCGATTCAGTGATGTTC 35771


RESULT   9
AL353791/c
LOCUS       AL353791   208763 bp    DNA              HTG       12-MAR-2001
DEFINITION  Homo sapiens chromosome 9 clone RP11-475B17, *** SEQUENCING IN
            PROGRESS ***, 4 unordered pieces.

ACCESSION   AL353791
VERSION     AL353791.5  GI:13373919
```

```
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 208763)
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 16, 2001 this sequence version replaced gi:11340258.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA472F14
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 208161 bases at least Q40
            Consensus quality: 208281 bases at least Q30
            Consensus quality: 208368 bases at least Q20
            Insert size: 208463; sum-of-contigs
            Insert size: 209014; 4.8% error; agarose-fp
            Quality coverage: 11.65x in Q20 bases; sum-of-contigs Quality
            coverage: 11.81x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1   119577: contig of 119577 bp in length
            *   119578 119677: gap of      100 bp
            *   119678   144219: contig of 24542 bp in length
            *   144220 144319: gap of      100 bp
            *   144320   204287: contig of 59968 bp in length
            *   204288 204387: gap of      100 bp
            *   204388   208763: contig of 4376 bp in length.
FEATURES             Location/Qualifiers
     source          1. .208763
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-475B17"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .119577
                     /note="assembly_fragment:00460
                     fragment_chain:1"
     misc_feature    119678. .144219
                     /note="assembly_fragment:00991
                     fragment_chain:1"
     misc_feature    144320. .204287
                     /note="assembly_fragment:04392
                     fragment_chain:1"
     misc_feature    204388. .208763
                     /note="assembly_fragment:05279.0"
BASE COUNT    60862 a  42636 c  42179 g  62780 t    306 others
ORIGIN
```

Query Match 15.4%; Score 93.6; DB 79; Length 208763;
Best Local Similarity 66.2%; Pred. No. 1.2e-16;
Matches 135; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```
Qy      30 gaatatttgcaattcctaaagacgtgttttgattgtcacacctgggtggggaacatgcta 89
           ||   |||||||| ||  || | | ||  | |  |   |  |||  |||   |||||
Db 146724 GACATTTTGCAATATCTGGAGGCATTTTCAGTTATTATAAGTGGGGCTGGGTGGATGCTG 146665

Qy      90 ctggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacgcacaggacag 149
           ||||||||| ||| | ||||  |||  |||||| |||||||| |  || ||||||||||
Db 146664 CTGGCATCTCATGGAGAGAGACCAGAGATGCTGTTAAACATCCTATAATGCACAGGACAG 146605

Qy     150 agccccacaaaagagaattatctagcccccaaatgtccataacactgctgttgagaaaacc 209
           |||||||| | |||||||||  ||||  |||||||  ||   |||  | |||||||  |
Db 146604 GCCCCCACAACACAGAATTATCCAGCCTAAAATGTCACTAGTGCTGAGGGTGAGAAACTC 146545

Qy     210 taccgcaggatcttactgggcttc 233
           |  | ||     | | ||   |||
Db 146544 TGACACATCGATTCAGTGATGTTC 146521
```

```
RESULT  10
AC016951
LOCUS       AC016951   147929 bp    DNA              PRI       05-MAY-2000
DEFINITION  Homo sapiens 3 BAC RP11-452H12 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC016951
VERSION     AC016951.9  GI:7658300
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 147929)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Garcia,D.K., Gorrell,J.H.,
            Gorrell,L.L., Guevara,W., Harris,K., He,X., Hernandez,J.,
            Hodgson,A., Hogues,M., Holloway,C., Hosak,H., Jackson,L.E.,
            Jackson,L., Jia,Y., Jones,M., Kelly,S., Kondejewski,N., Kong,Y.,
            Kovar,C., Leal,B., Li,Z., Lichtarge,O., Liu,J., Liu,W., Logan,O.,
            Lozado,R.J., Lu,J., Lucier,R., Martin,R., Martinez,C., McLeod,M.P.,
            Mei,G., Moore,S., Moorish,T., Morgan,M., Morris,S., Nash,S.,
            Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B.,
            Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D.,
            Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H.,
            Simon,M., Sparks,A., Stamps,A., Sucgang,R., Tabor,P., Taylor,T.,
            Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S.,
            Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,
            Wrensford,G., Yu,W., Zhou,X., Naylor,S.L., Nelson,D. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 147929)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 147929)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 147929)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 28, 2000 this sequence version replaced gi:7331304.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
```

entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT----------

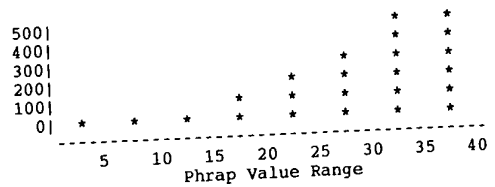
---------- Summary Statistics ----------

Contig length:	147929
Phrap values in estimate:	146381
Average error rate (BCM-Phrap estimate):	0.000137522
Fraction of Phrap values less than 40 :	0.0255019
Number of consensus changing edits:	17
Number of N's in consensus :	0

---------- Consensus changing edits ----------

Position	Original+Context	Edited+Context
26320	acctaatgct(n)gatgacgagt	acctaatgct(a)gatgacgagt
31259	ttctttttgg(n)aagatgaaac	ttctttttgg(g)aagatgaaac
37952	gaatatattt(n)gaaataagat	gaatatattt(g)gaaataagat
48020	agattcataa(n)gcaagtcctg	agattcataa(a)gcaagtcctg
61999	ttctaaaatt(n)gttttgactt	ttctaaaatt(a)gttttgactt
72566	tgattttgtt(n)ccttggaagc	tgattttgtt(t)ccttggaagc
81082	ctcatatagc(n)tacatactag	ctcatatagc(c)tacatactag
81162	atgaagtaag(n)caaagccatg	atgaagtaag(a)caaagccatg
81177	gccatggagg(n)agatgggagg	gccatggagg(g)agatgggagg
82666	aaattggaaa(n)aatccacact	aaattggaaa(c)aatccacact
82977	atgatagtgg(n)tacacaaatc	atgatagtgg(t)tacacaaatc
91636	ccgtccatgc(n)caaaaaaaaa	ccgtccatgc(a)aaaaaaaaaa
91637	cgtccatgcn(c)aaaaaaaaaa	cgtccatgca(a)aaaaaaaaaa
96541	aaaaaaaatt(t)aaaaaaaaaa	aaaaaaaatt(a)aaaaaaaaaa
96574	aaaaaaaaag(g)aaggatgtca	aaaaaaaaag(a)aaggatgtca
98037	gctgcaagca(n)gccgggaagc	gctgcaagca(t)gccgggaagc
139118	gaggcagagg(n)caggaaagga	gaggcagagg(c)caggaaagga

--------- Distribution of Quality < 40 Bases ----------



Version: 1.01 qxfo.

FEATURES	Location/Qualifiers
source	1. .147929
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="3"
	/clone="RP11-452H12"
repeat_region	complement(532. .617)
	/rpt_family="MSTB"
repeat_region	625. .696
	/rpt_family="LTR9"
STS	668. .891
	/standard_name="Z53523"
	/db_xref="dbSTS:45636"
repeat_region	697. .731
	/rpt_family="(CA)n"
repeat_region	complement(732. .1090)
	/rpt_family="MSTB"
repeat_region	1927. .2384
	/rpt_family="MLT1G1"
repeat_region	3079. .3391
	/rpt_family="AluJb"
repeat_region	complement(3685. .3897)
	/rpt_family="MER20"
repeat_region	complement(5392. .5890)
	/rpt_family="Charliela"
repeat_region	complement(5935. .6618)
	/rpt_family="Charliela"
STS	8265. .8469
	/standard_name="G09350"
	/db_xref="dbSTS:15731"
repeat_region	8348. .8390
	/rpt_family="(TTA)n"
repeat_region	8503. .8814
	/rpt_family="AluSx"
repeat_region	9081. .9226
	/rpt_family="MER53"
repeat_region	9245. .9413
	/rpt_family="MSTD"
repeat_region	9414. .9705
	/rpt_family="AluJb"
repeat_region	9706. .9951
	/rpt_family="MSTD"
repeat_region	9952. .10035
	/rpt_family="MER53"
repeat_region	11657. .13087
	/rpt_family="pTR5"
repeat_region	complement(13237. .13437)
	/rpt_family="MLT1H"
repeat_region	19462. .19885
	/rpt_family="HAL1"
repeat_region	20035. .20336
	/rpt_family="AluJb"
repeat_region	20395. .21174
	/rpt_family="L1PA2"
repeat_region	21170. .26420
	/rpt_family="L1PA2"
STS	25950. .26082
	/standard_name="G19948"
	/db_xref="dbSTS:32826"

```
 Query Match           15.1%;  Score 91.8;  DB 87;  Length 147929;
 Best Local Similarity  64.2%;  Pred. No. 4e-16;
 Matches  138;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy      2 gagtttgccccttacccccatcccagtgaatatttgcaattcctaaagacgtgtttga 61
          || | || | ||   |||| | |||||  |||||||   | |           ||| |
Db  28127 GATTCTGGCACCAACCCCCAACCCCAGGAAATATTTAACAATTTCTGGAGATATTTTGGC 28186

Qy     62 ttgtcacacctgggtggggaacatgctactggcatctaatgcatagagggcagtaatgct 121
          |      || || |  | |    ||||||||||||||||||| |||||| ||   || ||
Db  28187 TGTCACAACTTGAGGAGAGGGGGTGCTACTGGCATCTAATGAATAGAGGCCAAGGATCCT 28246

Qy    122 gctaaacatctttcaacgcacaggacagagccccacaaaagagaattatctagccccaaa 181
          ||||||  |  | ||| ||||||||||  |||||||| | ||||||||  || || |||
Db  28247 GCTAAATCTTCTACAATGCACAGGACAGCTCCCCACAACAAAGAATTATTCAGACCAAAA 28306

Qy    182 tgtccataacactgctgttgagaaaacctaccgca 216
          | ||  |||  | |  | |||||||| || | | |
Db  28307 TTTCACTAATGCAGAGGATGAGAAAAACTGCTGTA 28341


RESULT  11
AL353729
LOCUS       AL353729   194313 bp    DNA             HTG       06-MAR-2001
DEFINITION  Homo sapiens chromosome 9 clone RP11-290L7, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL353729
VERSION     AL353729.7  GI:13273646
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 194313)
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 11, 2001 this sequence version replaced gi:9800794.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA290L7
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 193402 bases at least Q40
            Consensus quality: 193962 bases at least Q30
            Consensus quality: 194209 bases at least Q20
            Insert size: 194313; sum-of-contigs
            Insert size: 196358; 3.1% error; agarose-fp
            Quality coverage: 6.31x in Q20 bases; sum-of-contigs Quality
            coverage: 6.45x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .194313
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-290L7"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .194313
                     /note="assembly_fragment:01008
                     clone_end:SP6
                     vector_side:left"
BASE COUNT    59225 a  36880 c  38036 g  60172 t
ORIGIN


 Query Match           15.0%;  Score 91;  DB 79;  Length 194313;
 Best Local Similarity  67.9%;  Pred. No. 7.1e-16;
 Matches  127;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy     30 gaatatttgcaattcctaaagacgtgttttgattgtcacacctgggtggggaacatgcta 89
          ||   |||||||| ||  || | | ||  | |  |   |  |||  |||   ||||
Db 167319 GACATTTTGCAATATCTGGAGGCATCTTCAGTTATTATAAGTGGGGCTGGGTGGGTGCTG 167378

Qy     90 ctggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacgcacaggacag 149
          |||||||||| ||| | ||||  |||  |||||| |||||||| |  || ||||||||||
Db 167379 CTGGCATCTCATGGAGAGAGACCAGAGATGCTGTTAAACATCCTATAATGCACAGGACAG 167438

Qy    150 agccccacaaaagagaattatctagccccaaatgtccataacactgctgttgagaaaacc 209
          ||||||||| | |||||||||| ||||  |||||||  ||   |||  | |||||||  |
Db 167439 GCCCCCACAACACAGAATTATCCAGCCTAAAATGTCACTAGTGCTGAGGGTGAGAAACTC 167498

Qy    210 taccgca 216
          |  | ||
Db 167499 TGACACA 167505


RESULT  12
AC017108/c
LOCUS       AC017108   207746 bp    DNA             HTG       02-SEP-2000
DEFINITION  Homo sapiens chromosome UL clone RP11-566J10, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
ACCESSION   AC017108
VERSION     AC017108.4  GI:9965023
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 207746)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 207746)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 2, 2000 this sequence version replaced gi:9838278.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0566J10
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 83% of reads
            Chemistry: Dye-terminator Big Dye; 17% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 191438 bases at least Q40
            Consensus quality: 197227 bases at least Q30
            Consensus quality: 200927 bases at least Q20
            Insert size: 194000; agarose-fp
            Insert size: 206046; sum-of-contigs
            Quality coverage: 4.36 in Q20 bases; agarose-fp
            Quality coverage: 3.63 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
```

```
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     3206: contig of 3206 bp in length
            *    3207     3306: gap of unknown length
            *    3307    11757: contig of 8451 bp in length
            *   11758    11857: gap of unknown length
            *   11858    24934: contig of 13077 bp in length
            *   24935    25034: gap of unknown length
            *   25035    36571: contig of 11537 bp in length
            *   36572    36671: gap of unknown length
            *   36672    50720: contig of 14049 bp in length
            *   50721    50820: gap of unknown length
            *   50821    61664: contig of 10844 bp in length
            *   61665    61764: gap of unknown length
            *   61765    77240: contig of 15476 bp in length
            *   77241    77340: gap of unknown length
            *   77341   104996: contig of 27656 bp in length
            *  104997   105096: gap of unknown length
            *  105097   132763: contig of 27667 bp in length
            *  132764   132863: gap of unknown length
            *  132864   160180: contig of 27317 bp in length
            *  160181   160280: gap of unknown length
            *  160281   164922: contig of 4642 bp in length
            *  164923   165022: gap of unknown length
            *  165023   166927: contig of 1905 bp in length
            *  166928   167027: gap of unknown length
            *  167028   170224: contig of 3197 bp in length
            *  170225   170324: gap of unknown length
            *  170325   173939: contig of 3615 bp in length
            *  173940   174039: gap of unknown length
            *  174040   181101: contig of 7062 bp in length
            *  181102   181201: gap of unknown length
            *  181202   189379: contig of 8178 bp in length
            *  189380   189479: gap of unknown length
            *  189480   199185: contig of 9706 bp in length
            *  199186   199285: gap of unknown length
            *  199286   207746: contig of 8461 bp in length.
FEATURES             Location/Qualifiers
     source          1. .207746
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="UL"
                     /clone="RP11-566J10"
     misc_feature    1. .3206
                     /note="assembly_name:Contig1"
     misc_feature    3307. .11757
                     /note="assembly_name:Contig10"
     misc_feature    11858. .24934
                     /note="assembly_name:Contig11"
     misc_feature    25035. .36571
                     /note="assembly_name:Contig12
                     clone_end:T7
                     vector_side:right"
     misc_feature    36672. .50720
                     /note="assembly_name:Contig13"
     misc_feature    50821. .61664
                     /note="assembly_name:Contig14
                     clone_end:SP6
                     vector_side:left"
     misc_feature    61765. .77240
                     /note="assembly_name:Contig15"
     misc_feature    77341. .104996
                     /note="assembly_name:Contig16"
     misc_feature    105097. .132763
                     /note="assembly_name:Contig17"
     misc_feature    132864. .160180
                     /note="assembly_name:Contig18"
     misc_feature    160281. .164922
                     /note="assembly_name:Contig2"
     misc_feature    165023  .166927
                     /note="assembly_name:Contig3"
     misc_feature    167028. .170224
                     /note="assembly_name:Contig4"
     misc_feature    170325. .173939
                     /note="assembly_name:Contig5"
     misc_feature    174040. .181101
                     /note="assembly_name:Contig6"
     misc_feature    181202. .189379
                     /note="assembly_name:Contig7"
     misc_feature    189480. .199185
                     /note="assembly_name:Contig8"
     misc_feature    199286. .207746
                     /note="assembly_name:Contig9"
BASE COUNT    60979 a  43162 c  40763 g  61116 t   1726 others
ORIGIN


  Query Match           15.0%;  Score 91;  DB 64;  Length 207746;
  Best Local Similarity  66.3%;  Pred. No. 7.1e-16;
  Matches  130;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy      24 cccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctgggtggggaac 83
           ||||| | | || ||  | | |          ||| | |||||||| || || || |
Db   87414 CCCAGGGNACATGTGGCAATGTTTGGAGACACTTTAGGTTGTCACAACTTGGGGGAGGTG 87355

Qy      84 atgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacgcaca 143
           |||||||||||| || |   |||||| |||  |  |||||||||||| | ||| |||||
Db   87354 GTGCTACTGGCATTTAGTAGGTAGAGGTCAGAGACACTGCTAAACATCCTGCAATGCACA 87295

Qy     144 ggacagagccccacaaaagagaattatctagccccaaatgtccataacactgctgttgag 203
           |||||||  |||||||  |||||||||  | |  ||||| | ||    |||    ||||
Db   87294 GGACAGACACCCACAACAAAGAATTATCCTGTCTAAAATGACAATTGTGCTGAGACTGAG 87235

Qy     204 aaaacctaccgcagga 219
           ||  ||| |   ||||
Db   87234 AAGCCCTGCTTTAGGA 87219


RESULT  13
AC021051
LOCUS       AC021051   162082 bp    DNA             HTG       04-NOV-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-214M2, WORKING DRAFT SEQUENCE,
            6 unordered pieces.
ACCESSION   AC021051
VERSION     AC021051.9  GI:11079339
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162082)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
```

Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 162082)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (13-JAN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 3, 2000 this sequence version replaced gi:11067234.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: HMZR
Center clone name: RP11-214M2
---------- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 144161 bases at least Q40
Consensus quality: 155224 bases at least Q30
Consensus quality: 158730 bases at least Q20
Estimated insert size: 160274; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
---
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 44978: contig of 44978 bp in length
* 44979 45078: gap of unknown length
* 45079 82822: contig of 37744 bp in length
* 82823 82922: gap of unknown length
* 82923 114983: contig of 32061 bp in length
* 114984 115083: gap of unknown length
* 115084 147193: contig of 32110 bp in length
* 147194 147293: gap of unknown length
* 147294 159999: contig of 12706 bp in length
* 160000 160099: gap of unknown length
* 160100 162082: contig of 1983 bp in length.

FEATURES Location/Qualifiers
source 1. .162082
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-214M2"

BASE COUNT 43856 a 35586 c 36576 g 45519 t 545 others

ORIGIN

Query Match 15.0%; Score 90.8; DB 66; Length 162082;
Best Local Similarity 64.0%; Pred. No. 8e-16;
Matches 137; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy       3 agtttgccccctttaccccccatcccagtgaatatttgcaattcctaaagacgtgttttgat 62
          | | || | ||   |||| | |||||  |||||||   | |           ||| | |
Db  119366 ATTCTGGCACCAACCCCCAACCCCAGGAAATATTTAACAATTTCTGGAGATATTTTGGCT 119425

Qy      63 tgtcacacctgggtggggaacatgctactggcatctaatgcatagagggcagtaatgctg 122
             || || |  | |    ||||||||||||||||||| |||||||  ||   || |||
Db  119426 GTCACAACTTGAGGAGAGGGGGTGCTACTGGCATCTAATGAATAGAGGCCAAGGATCCTG 119485

Qy     123 ctaaacatctttcaacgcacaggacagagccccacaaaagagaattatctagccccaaat 182
           |||||  |  | ||| ||||||||||  |||||||| | ||||||||  || || ||||
Db  119486 CTAAATCTTCTACAATGCACAGGACAGCTCCCCACAACAAAGAATTATTCAGACCAAAAT 119545

Qy     183 gtccataacactgctgttgagaaaacctaccgca 216
            ||  |||  | |  | |||||||| || | | |
Db  119546 TTCAGTAATGCAGAGGATGAGAAAAACTGCTGTA 119579
```

RESULT 14
AC073244
LOCUS AC073244 180861 bp DNA HTG 10-NOV-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-207J1, WORKING DRAFT SEQUENCE, 28 unordered pieces.
ACCESSION AC073244
VERSION AC073244.7 GI:11128176
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 180861)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

```
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
            and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180861)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA

COMMENT     On Nov 9, 2000 this sequence version replaced gi:9966645.
            ---------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
              Center project name: HBLQ
              Center clone name: RP11-207J1
            ---------- Summary Statistics
              Sequencing vector: M13; L08821
              Chemistry: Dye-terminator Big Dye: 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 154874 bases at least Q40
              Consensus quality: 166894 bases at least Q30
              Consensus quality: 172186 bases at least Q20
              Estimated insert size: 172998; sum-of-contigs estimation
              Quality coverage: 0x in Q20 bases; agarose-fp estimation
              Quality coverage: 3.2x in Q20 bases; sum-of-contigs estimation
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 28 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    34187: contig of 34187 bp in length
            *    34188    34287: gap of unknown length
            *    34288    55848: contig of 21561 bp in length
            *    55849    55948: gap of unknown length
            *    55949    71980: contig of 16032 bp in length
            *    71981    72080: gap of unknown length
            *    72081    84930: contig of 12850 bp in length
            *    84931    85030: gap of unknown length
            *    85031    94603: contig of 9573 bp in length
            *    94604    94703: gap of unknown length
            *    94704   102324: contig of 7621 bp in length
            *   102325   102424: gap of unknown length
            *   102425   109333: contig of 6909 bp in length
            *   109334   109433: gap of unknown length
            *   109434   116085: contig of 6652 bp in length
            *   116086   116185: gap of unknown length
            *   116186   123583: contig of 7398 bp in length
            *   123584   123683: gap of unknown length
            *   123684   130880: contig of 7197 bp in length
            *   130881   130980: gap of unknown length
            *   130981   135776: contig of 4796 bp in length
            *   135777   135876: gap of unknown length
            *   135877   139816: contig of 3940 bp in length
            *   139817   139916: gap of unknown length
            *   139917   145457: contig of 5541 bp in length
            *   145458   145557: gap of unknown length
            *   145558   148749: contig of 3192 bp in length
            *   148750   148849: gap of unknown length
            *   148850   152774: contig of 3925 bp in length
            *   152775   152874: gap of unknown length
            *   152875   156616: contig of 3742 bp in length
            *   156617   156716: gap of unknown length
            *   156717   160503: contig of 3787 bp in length
            *   160504   160603: gap of unknown length
            *   160604   164046: contig of 3443 bp in length
            *   164047   164146: gap of unknown length
            *   164147   165789: contig of 1643 bp in length
            *   165790   165889: gap of unknown length
            *   165890   168039: contig of 2150 bp in length
            *   168040   168139: gap of unknown length
            *   168140   170342: contig of 2203 bp in length
            *   170343   170442: gap of unknown length
            *   170443   172779: contig of 2337 bp in length
            *   172780   172879: gap of unknown length
            *   172880   174521: contig of 1642 bp in length
            *   174522   174621: gap of unknown length
            *   174622   175860: contig of 1239 bp in length
            *   175861   175960: gap of unknown length
            *   175961   176987: contig of 1027 bp in length
            *   176988   177087: gap of unknown length
            *   177088   178335: contig of 1248 bp in length
            *   178336   178435: gap of unknown length
            *   178436   179723: contig of 1288 bp in length
            *   179724   179823: gap of unknown length
            *   179824   180861: contig of 1038 bp in length.
FEATURES             Location/Qualifiers
     source          1. .180861
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-207J1"
BASE COUNT    51233 a  39202 c  39227 g  48472 t   2727 others
ORIGIN


  Query Match           15.0%;  Score 90.8;  DB 74;  Length 180861;
  Best Local Similarity  64.0%;  Pred. No. 8e-16;
  Matches  137;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy       3 agtttgccccctttaccccatcccagtgaatatttgcaattcctaaagacgtgttttgat 62
           | | || | ||   |||| | |||||  |||||||   | |                ||| | |
Db   45482 ATTCTGGCACCAACCCCCAACCCCAGGAAATATTTAACAATTTCTGGAGATATTTTGGCT 45541

Qy      63 tgtcacacctgggtggggaacatgctactggcatctaatgcatagagggcagtaatgctg 122
             || || |  | |    ||||||||||||||||||| |||||||  ||    || |||
Db   45542 GTCACAACTTGAGGAGAGGGGGTGCTACTGGCATCTAATGAATAGAGGCCAAGGATCCTG 45601

Qy     123 ctaaacatctttcaacgcacaggacagagccccacaaaagagaattatctagccccaaat 182
           |||||  |  | ||| ||||||||||  |||||||| | ||||||||  || || ||||
Db   45602 CTAAATCTTCTACAATGCACAGGACAGCTCCCCACAACAAAGAATTATTCAGACCAAAAT 45661

Qy     183 gtccataacactgctgttgagaaaacctaccgca 216
             ||  |||  | |  | |||||||| || | | |
Db   45662 TTCAGTAATGCAGAGGATGAGAAAAACTGCTGTA 45695


RESULT  15
AC022612/c
LOCUS       AC022612    63020 bp    DNA             HTG       18-MAR-2001
DEFINITION  Homo sapiens clone RP11-659D6, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC022612
VERSION     AC022612.3  GI:13376927
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 63020)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-659D6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 63020)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
```

Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Mar 18, 2001 this sequence version replaced gi:9127546.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L4317
Center clone name: 659_D_6
----------.
* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 705: contig of 705 bp in length
* 706 805: gap of 100 bp
* 806 1565: contig of 760 bp in length
* 1566 1665: gap of 100 bp
* 1666 2393: contig of 728 bp in length
* 2394 2493: gap of 100 bp
* 2494 3185: contig of 692 bp in length
* 3186 3285: gap of 100 bp
* 3286 4007: contig of 722 bp in length
* 4008 4107: gap of 100 bp
* 4108 4798: contig of 691 bp in length
* 4799 4898: gap of 100 bp
* 4899 5682: contig of 784 bp in length
* 5683 5782: gap of 100 bp
* 5783 6491: contig of 709 bp in length
* 6492 6591: gap of 100 bp
* 6592 7280: contig of 689 bp in length
* 7281 7380: gap of 100 bp
* 7381 8087: contig of 707 bp in length
* 8088 8187: gap of 100 bp
* 8188 8895: contig of 708 bp in length
* 8896 8995: gap of 100 bp
* 8996 9724: contig of 729 bp in length
* 9725 9824: gap of 100 bp
* 9825 10558: contig of 734 bp in length
* 10559 10658: gap of 100 bp
* 10659 11337: contig of 679 bp in length
* 11338 11437: gap of 100 bp
* 11438 12151: contig of 714 bp in length
* 12152 12251: gap of 100 bp
* 12252 12959: contig of 708 bp in length
* 12960 13059 gap of 100 bp
* 13060 13753: contig of 694 bp in length
* 13754 13853: gap of 100 bp
* 13854 14552: contig of 699 bp in length
* 14553 14652: gap of 100 bp
* 14653 15362: contig of 710 bp in length
* 15363 15462: gap of 100 bp
* 15463 16165: contig of 703 bp in length
* 16166 16265: gap of 100 bp
* 16266 16968: contig of 703 bp in length
* 16969 17068: gap of 100 bp
* 17069 17792: contig of 724 bp in length
* 17793 17892: gap of 100 bp
* 17893 18601: contig of 709 bp in length
* 18602 18701: gap of 100 bp
* 18702 19397: contig of 696 bp in length
* 19398 19497: gap of 100 bp
* 19498 20203: contig of 706 bp in length
* 20204 20303: gap of 100 bp
* 20304 20994: contig of 691 bp in length
* 20995 21094: gap of 100 bp
* 21095 21795: contig of 701 bp in length
* 21796 21895: gap of 100 bp
* 21896 22600: contig of 705 bp in length
* 22601 22700: gap of 100 bp
* 22701 23379: contig of 679 bp in length
* 23380 23479: gap of 100 bp
* 23480 24186: contig of 707 bp in length
* 24187 24286: gap of 100 bp
* 24287 24999: contig of 713 bp in length
* 25000 25099: gap of 100 bp
* 25100 25860: contig of 761 bp in length
* 25861 25960: gap of 100 bp
* 25961 26662: contig of 702 bp in length
* 26663 26762: gap of 100 bp
* 26763 27471: contig of 709 bp in length
* 27472 27571: gap of 100 bp
* 27572 28264: contig of 693 bp in length
* 28265 28364: gap of 100 bp
* 28365 29118: contig of 754 bp in length
* 29119 29218: gap of 100 bp
* 29219 29884: contig of 666 bp in length
* 29885 29984: gap of 100 bp
* 29985 30669: contig of 685 bp in length
* 30670 30769: gap of 100 bp
* 30770 31518: contig of 749 bp in length
* 31519 31618: gap of 100 bp
* 31619 32320: contig of 702 bp in length
* 32321 32420: gap of 100 bp
* 32421 33139: contig of 719 bp in length
* 33140 33239: gap of 100 bp
* 33240 34000: contig of 761 bp in length
* 34001 34100: gap of 100 bp
* 34101 34826: contig of 726 bp in length
* 34827 34926: gap of 100 bp
* 34927 35629: contig of 703 bp in length
* 35630 35729: gap of 100 bp
* 35730 36444: contig of 715 bp in length
* 36445 36544: gap of 100 bp
* 36545 37263: contig of 719 bp in length
* 37264 37363: gap of 100 bp
* 37364 38044: contig of 681 bp in length
* 38045 38144: gap of 100 bp
* 38145 38838: contig of 694 bp in length
* 38839 38938: gap of 100 bp
* 38939 39631: contig of 693 bp in length
* 39632 39731: gap of 100 bp
* 39732 40471: contig of 740 bp in length
* 40472 40571: gap of 100 bp
* 40572 41271: contig of 700 bp in length
* 41272 41371: gap of 100 bp
* 41372 42084: contig of 713 bp in length
* 42085 42184: gap of 100 bp
* 42185 42883: contig of 699 bp in length

```
          *    42884 42983: gap of      100 bp
          *    42984    43690: contig of 707 bp in length
          *    43691 43790: gap of      100 bp
          *    43791    44484: contig of 694 bp in length
          *    44485 44584: gap of      100 bp
          *    44585    45265: contig of 681 bp in length
          *    45266 45365: gap of      100 bp
          *    45366    46078: contig of 713 bp in length
          *    46079 46178: gap of      100 bp
          *    46179    46886: contig of 708 bp in length
          *    46887 46986: gap of      100 bp
          *    46987    47700: contig of 714 bp in length
          *    47701 47800: gap of      100 bp
          *    47801    48505: contig of 705 bp in length
          *    48506 48605: gap of      100 bp
          *    48606    49293: contig of 688 bp in length
          *    49294 49393: gap of      100 bp
          *    49394    50141: contig of 748 bp in length
          *    50142 50241: gap of      100 bp
          *    50242    50943: contig of 702 bp in length
          *    50944 51043: gap of      100 bp
          *    51044    51745: contig of 702 bp in length
          *    51746 51845: gap of      100 bp
          *    51846    52554: contig of 709 bp in length
          *    52555 52654: gap of      100 bp
          *    52655    53348: contig of 694 bp in length
          *    53349 53448: gap of      100 bp
          *    53449    54172: contig of 724 bp in length
          *    54173 54272: gap of      100 bp
          *    54273    54957: contig of 685 bp in length
          *    54958 55057: gap of      100 bp
          *    55058    55758: contig of 701 bp in length
          *    55759 55858: gap of      100 bp
          *    55859    56570: contig of 712 bp in length
          *    56571 56670: gap of      100 bp
          *    56671    57378: contig of 708 bp in length
          *    57379 57478: gap of      100 bp

 Query Match             14.9%;  Score 90.4;  DB 67;  Length 63020;
 Best Local Similarity   66.3%;  Pred. No. 9.5e-16;
 Matches  130;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy      24 cccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctgggtggggaac 83
           ||||| | | || ||  | |  |         ||| | |||||||| || || || |
Db   45912 CCCAGGGGACATGTGGCAATGTTTGGAGACACTTTAGGTTGTCACAACTTGGGGGAGGTG 45853

Qy      84 atgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacgcaca 143
           |||||||||||| || |   |||||| |||  |  |||||||||||| | ||| |||||
Db   45852 GTGCTACTGGCATTTAGTAGGTAGAGGTCAGAGACACTGCTAAACATCCTGCAATGCACA 45793

Qy     144 ggacagagccccacaaaagagaattatctagccccaaatgtccataacactgctgttgag 203
           |||||||  ||||||| | |||||||||  | |  |||||| | ||    |||    ||||
Db   45792 GGACAGACACCCACAACAAAGAATTATCCTGTCTAAAATGACAATTGTGCTGAGACTGAG 45733

Qy     204 aaaacctaccgcagga 219
           ||  ||| |   ||||
Db   45732 AAGCCCTGCTTTAGGA 45717


Search completed: November  5, 2001, 23:02:50
Job time: 2800 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: November 5, 2001, 22:57:50 ; Search time 1188.78 Seconds
(without alignments)
7884.937 Million cell updates/sec

Title: US-09-656-668-198
Perfect score: 606
Sequence: 1 tgagtttgccccttacccc.....aagcctgttcttgtctgcac 606

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43 em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
	606	100.0	606	10	AX093380	AX093380 Sequence
2	59	9.7	90497	92	HS732E4	AL008722 Human DNA
3	26	4.3	105137	93	HSJ104O17	AL080316 Human DNA
4	25	4.1	154969	62	AC012443	AC012443 Homo sapi
5	25	4.1	183893	77	AC090051	AC090051 Homo sapi
6	25	4.1	194294	76	AC079344	AC079344 Homo sapi
c 7	23	3.8	67084	78	AC090599	AC090599 Homo sapi
8	23	3.8	93959	92	HS106C24	Z83313 Human DNA s

c	9	23	3.8	103757	75	AC073940
	10	23	3.8	127590	85	AC002554
c	11	23	3.8	134394	73	AC067916
	12	23	3.8	151357	62	AC011573
	13	23	3.8	157402	71	AC034161
	14	23	3.8	159723	66	AC021862
	15	23	3.8	171589	75	AC078953
c	16	23	3.8	172214	64	AC016155
c	17	23	3.8	172588	60	AC008590
c	18	23	3.8	173808	71	AC036206
c	19	23	3.8	174128	89	AL137139
c	20	23	3.8	176092	73	AC068222
	21	23	3.8	178028	79	AL355338
	22	23	3.8	178328	65	AC019265
	23	23	3.8	180316	86	AC007773
	24	23	3.8	186292	72	AC064876
	25	23	3.8	186699	77	AC090015
	26	23	3.8	189079	72	AC053545
c	27	23	3.8	190009	78	AC090519
	28	23	3.8	191060	66	AC021468
c	29	23	3.8	199497	68	AC024534
	30	23	3.8	207221	62	AC012564
c	31	22	3.6	161837	86	AC006479
c	32	22	3.6	174380	64	AC016075
c	33	22	3.6	178042	68	AC023986
	34	22	3.6	194058	86	AC007564
	35	22	3.6	225635	60	AC007553
c	36	21	3.5	39540	85	AC000079
c	37	21	3.5	43461	6	CER03D7
c	38	21	3.5	43934	85	AC000068
	39	21	3.5	90590	61	AC010629
	40	21	3.5	130757	76	AC079584
	41	21	3.5	137658	92	HS1022J11
c	42	21	3.5	139457	80	AL357652
	43	21	3.5	140002	90	AL354985
c	44	21	3.5	150803	81	AL450398
	45	21	3.5	155970	73	AC068085

AC073940	Homo sapi
AC002554	Human Chr
AC067916	Homo sapi
AC011573	Homo sapi
AC034161	Homo sapi
AC021862	Homo sapi
AC078953	Homo sapi
AC016155	Homo sapi
AC008590	Homo sapi
AC036206	Homo sapi
AL137139	Human DNA
AC068222	Homo sapi
AL355338	Homo sapi
AC019265	Homo sapi
AC007773	Homo sapi
AC064876	Homo sapi
AC090015	Homo sapi
AC053545	Homo sapi
AC090519	Homo sapi
AC021468	Homo sapi
AC024534	Homo sapi
AC012564	Homo sapi
AC006479	Homo sapi
AC016075	Homo sapi
AC023986	Homo sapi
AC007564	Homo sapi
AC007553	Homo sapi
AC000079	Homo sapi
Z46828	Caenorhabdi
AC000068	Homo sapi
AC010629	Homo sapi
AC079584	Homo sapi
AL049765	Human DNA
AL357652	Homo sapi
AL354985	Human DNA
AL450398	Homo sapi
AC068085	Homo sapi

ALIGNMENTS

```
RESULT   1
AX093380
LOCUS       AX093380      606 bp    DNA              PAT       30-MAR-2001
DEFINITION  Sequence 198 from Patent WO0118046.
ACCESSION   AX093380
VERSION     AX093380.1  GI:13509828
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 606)
  AUTHORS   Xu,J. and Stolk,J.A.
  TITLE     Ovarian tumor sequences and methods of use therefor
  JOURNAL   Patent: WO 0118046-A 198 15-MAR-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .606
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      175 a    147 c    123 g    161 t
ORIGIN
```

Query Match 100.0%; Score 606; DB 10; Length 606;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 606; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 tgagtttgccccccttacccccatcccagtgaatatttgcaattcctaaagacgtgttttg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGAGTTTGCCCCCTTACCCCCATCCCAGTGAATATTTGCAATTCCTAAAGACGTGTTTTG 60

Qy     61 attgtcacacctgggtggggaacatgctactggcatctaatgcatagagggcagtaatgc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATTGTCACACCTGGGTGGGGAACATGCTACTGGCATCTAATGCATAGAGGGCAGTAATGC 120

Qy    121 tgctaaacatctttcaacgcacaggacagagccccacaaaagagaattatctagccccaa 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGCTAAACATCTTTCAACGCACAGGACAGAGCCCCACAAAAGAGAATTATCTAGCCCCAA 180

Qy    181 atgtccataacactgctgttgagaaaacctaccgcaggatcttactgggcttcataggta 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATGTCCATAACACTGCTGTTGAGAAAACCTACCGCAGGATCTTACTGGGCTTCATAGGTA 240

Qy    241 agcttgccttttgttctggcttctgtagatatataaaataaagacactgcccagtccctcc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGCTTGCCTTTGTTCTGGCTTCTGTAGATATATAAAATAAAGACACTGCCCAGTCCCTCC 300

Qy    301 ctcaacgtcccgagccagggctcaaggcaattccaataacagtagaatgaacactaaata 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTCAACGTCCCGAGCCAGGGCTCAAGGCAATTCCAATAACAGTAGAATGAACACTAAATA 360

Qy    361 ttgatttcaaaatctcagcaactagaagaatgaccaaccatcctggttggcctgggactg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTGATTTCAAAATCTCAGCAACTAGAAGAATGACCAACCATCCTGGTTGGCCTGGGACTG 420

Qy    421 tcctagttttagcattgaaagtttcaggttccaggaaagccctcaggcctgggctgctgg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCCTAGTTTTAGCATTGAAAGTTTCAGGTTCCAGGAAAGCCCTCAGGCCTGGGCTGCTGG 480

Qy    481 tcaccctagcagctgagggactcttcaatacagaattagtctttgtgcactggagatgaa 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TCACCCTAGCAGCTGAGGGACTCTTCAATACAGAATTAGTCTTTGTGCACTGGAGATGAA 540

Qy    541 tatactttaatttgtaacatgtgaaaacatctataaacatctactgaagcctgttcttgt 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TATACTTTAATTTGTAACATGTGAAAACATCTATAAACATCTACTGAAGCCTGTTCTTGT 600

Qy    601 ctgcac 606
          ||||||
Db    601 CTGCAC 606
```

```
RESULT   2
HS732E4
LOCUS       HS732E4     90497 bp    DNA              PRI       12-DEC-1999
DEFINITION  Human DNA sequence from clone CTA-732E4 on chromosome 22q12.1
            Contains ESTs, STSs and GSSs, complete sequence.
ACCESSION   AL008722
VERSION     AL008722.16  GI:5419644
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 90497)
  AUTHORS   Lloyd,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 8, 1999 this sequence version replaced gi:5304865.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
```

together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep CTA-732E4 is from the human BAC library described in U-J. Kim et al. (1996) Genomics 34, 213-218.
VECTOR: pBeloBAC11
This sequence is the entire insert of clone CTA-732E4 The true left end of clone RP11-541J16 is at 5510 in this sequence. The true right end of clone CTA-544A11 is at 41939 in this sequence.

FEATURES	Location/Qualifiers
source	1. .90497 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="22" /map="q12.1" /clone="CTA-732E4" /clone_lib="CIT978SK-A2"
repeat_region	2. .631 /note="L1M1 repeat: matches 994. .1640 of consensus"
repeat_region	632. .757 /note="AluJb repeat: matches 1. .126 of consensus"
repeat_region	758. .1385 /note="L1P repeat: matches 3548. .4173 of consensus"
repeat_region	1374. .1869 /note="L1P repeat: matches 3221. .3716 of consensus"
repeat_region	1952. .2076 /note="L1MD2 repeat: matches 5142. .5259 of consensus"
repeat_region	2760. .3005 /note="L1MD2 repeat: matches 5259. .6173 of consensus"
repeat_region	3418. .3477 /note="30 copies 2 mer ac 90 conserved"
repeat_region	3423. .3478 /note="14 copies 4 mer caca 94 conserved"
repeat_region	3670. .3742 /note="MER82 repeat: matches 580. .653 of consensus"
repeat_region	3763. .4128 /note="MER82 repeat: matches 1. .386 of consensus"
misc_feature	3879. .4353 /note="match: GSS: Em:AQ559531"
repeat_region	4364. .4569 /note="L2 repeat: matches 2535. .2750 of consensus"
repeat_region	4612. .4930 /note="AluJo repeat: matches 1. .307 of consensus"
repeat_region	4934. .5220 /note="AluSx repeat: matches 1. .287 of consensus"
repeat_region	5221. .5268 /note="16 copies 3 mer taa 97 conserved"
misc_feature	complement(5266. .5727) /note="match: GSS: Em:AQ028870"
misc_feature	complement(5282. .5740) /note="match: GSS: Em:AQ557814"
repeat_region	5417. .5569 /note="MIR repeat: matches 2. .167 of consensus"
repeat_region	5645. .5708 /note="32 copies 2 mer at 71 conserved"
misc_feature	6469. .6726 /note="match: STS: Em:G03694; match: STS: Em:G03694"
repeat_region	7573. .7749 /note="MIR repeat: matches 22. .207 of consensus"
repeat_region	10322. .16495 /note="L1PA2 repeat: matches 7. .6146 of consensus"
misc_feature	16265. .16787 /note="match: GSS: Em:AQ186300"
repeat_region	16926. .16967 /note="21 copies 2 mer aa 76 conserved"
repeat_region	17850. .17971 /note="AluJo repeat matches 19. .147 of consensus"
repeat_region	17972. .18285 /note="AluSg1 repeat: matches 1. .309 of consensus"
repeat_region	18286. .18472 /note="AluJo repeat: matches 147. .309 of consensus"
repeat_region	19710. .20005 /note="AluSx repeat: matches 1. .296 of consensus"
misc_feature	20789. .21262 /note="match: GSS: Em:B14443"
misc_feature	20811. .21058 /note="match: STS: Em:L04555"
repeat_region	21369. .21551 /note="MIR repeat: matches 2. .187 of consensus"
repeat_region	22494. .22697 /note="MIR repeat: matches 35. .236 of consensus"
repeat_region	23443. .23755 /note="AluSx repeat: matches 1. .312 of consensus"
repeat_region	24052. .24348 /note="AluSq repeat: matches 1. .295 of consensus"
repeat_region	25623. .25670 /note="24 copies 2 mer ac 77 conserved"
repeat_region	26003. .26192 /note="MIR repeat: matches 20. .216 of consensus"
repeat_region	27067. .27138 /note="Charlie4 repeat: matches 1902. .1961 of consensus"
repeat_region	27363. .27661 /note="AluYb8 repeat: matches 1. .306 of consensus"
repeat_region	28662. .28722 /note="L2 repeat: matches 2683. .2744 of consensus"
repeat_region	28758. .28797 /note="10 copies 4 mer tata 87 conserved"
repeat_region	28759. .28798 /note="20 copies 2 mer at 87 conserved"
repeat_region	29334. .29421 /note="22 copies 4 mer gaaa 65 conserved"
repeat_region	29805. .30155 /note="L1MB5 repeat: matches 4639. .4984 of consensus"
repeat_region	30150. .31160 /note="L1MB5 repeat: matches 5157. .6166 of consensus"
repeat_region	31294. .31449 /note="L1PA2 repeat: matches 5991. .6146 of consensus"
repeat_region	31544. .31681 /note="MIR repeat: matches 93. .254 of consensus"
repeat_region	31914. .32062 /note="MIR repeat: matches 101. .261 of consensus"
repeat_region	32553. .32814 /note="L2 repeat: matches 1072. .1341 of consensus"
repeat_region	32884. .32952 /note="MER45B repeat: matches 2. .76 of consensus"
repeat_region	33181. .33436 /note="MER45C repeat: matches 676. .952 of consensus"
repeat_region	33504. .33671 /note="MIR repeat: matches 4. .191 of consensus"
repeat_region	33771. .34072 /note="AluJo repeat: matches 1. .310 of consensus"
misc_feature	complement(35126. .35616) /note="match: GSS: Em:AQ407436"
repeat_region	37167. .38583 /note="TIGGER1 repeat: matches 2. .1394 of consensus"
repeat_region	38584. .38937 /note="THE1B repeat: matches 3. .364 of consensus"
repeat_region	38938. .40548 /note="THE1B-INTERNAL repeat: matches 1. .1580 of consensus"
repeat_region	40551. .40904 /note="THE1B repeat: matches 3. .364 of consensus"
repeat_region	40905. .41859 /note="TIGGER1 repeat: matches 1394. .2418 of consensus"
misc_feature	complement(41368. .41893) /note="match: GSS: Em:B18147"
misc_feature	complement(41562. .41939) /note="match: GSS: Em:AQ074435"
misc_feature	complement(41582. .41936) /note="match: GSS: Em:B14391; match: STS: Em:B14391"

```
     repeat_region   42857. .42971
                     /note="L2 repeat: matches 2540. .2659 of consensus"
     repeat_region   43009. .45826
                     /note="L1M1 repeat: matches 2469. .4924 of consensus"
     repeat_region   46433. .46883
                     /note="LTR33 repeat: matches 24. .472 of consensus"
     repeat_region   47129. .47240
                     /note="28 copies 4 mer gaga 66 conserved"
     repeat_region   47132. .47241
                     /note="55 copies 2 mer ag 67 conserved"
     repeat_region   47276. .47434
                     /note="MER5A repeat: matches 4. .189 of consensus"
     repeat_region   47698. .47991
                     /note="AluSc repeat: matches 1. .299 of consensus"
     repeat_region   48005. .48060
                     /note="FLAM_A repeat: matches 81. .133 of consensus"
     repeat_region   48061. .48341
                     /note="AluSg1 repeat: matches 1. .293 of consensus"
     repeat_region   48399. .48549
                     /note="FRAM repeat: matches -5. .146 of consensus"
     misc_feature    49002. .49450
                     /note="match: STS: Em:R42525; match: STS: Em:G24402"
     misc_feature    49916. .50351
                     /note="match: GSS: Em:AQ805065"
     misc_feature    50076. .50449

  Query Match             9.7%;  Score 59;  DB 92;  Length 90497;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-23;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     528 cactggagatgaatatactttaatttgtaacatgtgaaaacatctataaacatctactg 586
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20114 CACTGGAGATGAATATACTTTAATTTGTAACATGTGAAAACATCTATAAACATCTACTG 20172


RESULT   3
HSJ104O17
LOCUS       HSJ104O17  105137 bp    DNA              PRI       15-FEB-2000
DEFINITION  Human DNA sequence from clone RP1-104O17 on chromosome 6q16.1-16.3
            Contains STSs and GSSs, complete sequence.

ACCESSION   AL080316
VERSION     AL080316.8  GI:5650657
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1  (bases 1 to 105137)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk

COMMENT     On Jul 29, 1999 this sequence version replaced gi:5566542.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP1-104O17 is
            from the library RPCI-1 constructed at the Roswell Park Cancer
            Institute by the group of Pieter de Jong. For further details see
            http://bacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP1-104O17 This
            sequence was generated from part of bacterial clone contigs of
            human chromosome 6, constructed by the Sanger Centre Chromosome 6
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6.

FEATURES             Location/Qualifiers
     source          1. .105137
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="q16.1-16.3"
                     /clone_lib="RPCI-1"
                     /clone="RP1-104O17"
     misc_feature    complement(372. .811)
                     /note="match: GSS: Em:AQ006718"
     repeat_region   1037. .1080
                     /note="22 copies 2 mer ta 88% conserved"
     repeat_region   1153. .1456
                     /note="AluSx repeat: matches 1. .301 of consensus"
     repeat_region   2365. .2702
                     /note="L2 repeat: matches 2363. .2750 of consensus"
     repeat_region   3579. .3733
                     /note="L2 repeat: matches 2041. .2201 of consensus"
     repeat_region   4961. .5272
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   5545. .5648
                     /note="52 copies 2 mer ta 71% conserved"
     repeat_region   7388. .7668
                     /note="AluJo repeat: matches 2. .283 of consensus"
     repeat_region   7871. .8066
                     /note="MER58A repeat: matches 10. .224 of consensus"
     misc_feature    complement(10286. .10857)
                     /note="match: GSS: Em:B82911"
     repeat_region   11075. .11346
                     /note="L1ME3 repeat: matches 5657. .5940 of consensus"
     misc_feature    complement(13059. .13460)
                     /note="match: STS: Em:G21604"
     repeat_region   14002. .14199
                     /note="MER58C repeat: matches 3. .89 of consensus"
     repeat_region   17080. .17249
                     /note="L1MC1 repeat: matches 6165. .6323 of consensus"
     repeat_region   17425. .17538
                     /note="MIR repeat: matches 131. .245 of consensus"
     misc_feature    complement(17580. .17784)
                     /note="match: STS: Em:G43022"
     repeat_region   18332. .18525
                     /note="AluJo repeat: matches 87. .295 of consensus"
     repeat_region   18919. .19338
                     /note="L2 repeat: matches 1868. .2326 of consensus"
     repeat_region   19344. .19733
                     /note="MLT1B repeat: matches 12. .389 of consensus"
     repeat_region   20175. .20286
                     /note="56 copies 2 mer at 64% conserved"
     repeat_region   20512. .20976
                     /note="L2 repeat: matches 2264. .2750 of consensus"
     repeat_region   21113. .21217
                     /note="L2 repeat: matches 1233. .1341 of consensus"
     repeat_region   21331. .21517
                     /note="L1MEc repeat: matches 1768. .1953 of consensus"
     repeat_region   21882. .22020
                     /note="L1ME repeat: matches 419. .560 of consensus"
     repeat_region   22615. .22710
                     /note="48 copies 2 mer tt 62% conserved"
     repeat_region   23360. .23417
                     /note="29 copies 2 mer tt 69% conserved"
     repeat_region   24769. .25327
                     /note="HERV16 repeat: matches 6. .548 of consensus"
     repeat_region   25328. .26196
                     /note="MER6 repeat: matches 1. .864 of consensus"
     repeat_region   26197. .26818
                     /note="HERV16 repeat: matches 548. .1219 of consensus"
```

```
repeat_region   26964. .27296
                /note="LTR16A repeat: matches 95. .450 of consensus"
repeat_region   27297. .27643
                /note="HERVL repeat: matches 3587. .3946 of consensus"
repeat_region   28215. .28338
                /note="MSTA repeat: matches 284. .422 of consensus"
misc_feature    28589. .29009
                /note="match: GSS: Em:AQ308663"
repeat_region   28617. .28812
                /note="MSTA repeat: matches 1. .178 of consensus"
repeat_region   29010. .29383
                /note="MLT1B repeat: matches 1. .388 of consensus"
repeat_region   30078. .30297
                /note="LTR16A repeat: matches 15. .229 of consensus"
repeat_region   32881. .33162
                /note="L1MC5 repeat: matches 7649. .7926 of consensus"
repeat_region   33433. .33616
                /note="MIR repeat: matches 40. .260 of consensus"
misc_feature    complement(33617. .33834)
                /note="match: GSS: Em:AQ240281"
repeat_region   33899. .34036
                /note="MLT1I repeat: matches 269. .404 of consensus"
repeat_region   37901. .38084
                /note="AluSg/x repeat: matches 83. .264 of consensus"
repeat_region   38096. .38119
                /note="12 copies 2 mer ac 95% conserved"
repeat_region   38855. .38952
                /note="MIR repeat: matches 85. .184 of consensus"
repeat_region   39820. .40101
                /note="AluY repeat: matches 1. .299 of consensus"
repeat_region   40284. .41588
                /note="L1PA7 repeat: matches 2138. .3447 of consensus"
repeat_region   41588. .44294
                /note="L1PA7 repeat: matches 3441. .6142 of consensus"
repeat_region   45641. .46728
                /note="L1ME1 repeat: matches 4994. .6109 of consensus"
repeat_region   46737. .47808
                /note="TIGGER1 repeat: matches 1350. .2418 of consensus"
repeat_region   47815. .48092
                /note="Charlie1b repeat: matches 70. .386 of consensus"
repeat_region   48179. .48335
                /note="L1PB1 repeat: matches 5999. .6155 of consensus"
repeat_region   48338. .48369
                /note="16 copies 2 mer aa 93% conserved"
repeat_region   48370. .48882
                /note="L1M4 repeat: matches 4272. .4790 of consensus"
repeat_region   48927. .49175
                /note="L1MEc repeat: matches 2146. .2412 of consensus"
repeat_region   49180. .49590
                /note="L1MD3 repeat: matches 7310. .7739 of consensus"
repeat_region   49630. .49717
                /note="L1MC4 repeat: matches 7219. .7313 of consensus"
repeat_region   49732. .50676
                /note="L1MEc repeat: matches 1514. .2152 of consensus"
repeat_region   50716. .50836
                /note="L1MEc repeat: matches 1287. .1413 of consensus"
repeat_region   50849. .51871
                /note="L1MC1 repeat: matches 5287. .6316 of consensus"
repeat_region   51874. .52239
                /note="L1MEc repeat: matches 929. .1302 of consensus"
repeat_region   52268. .52519
                /note="L1MEc repeat: matches 523. .780 of consensus"
repeat_region   53098. .53206
                /note="Alu repeat: matches 197. .305 of consensus"
repeat_region   53208. .53388
                /note="MIR repeat: matches 41. .231 of consensus"
repeat_region   53414. .53830
                /note="L1PA8 repeat: matches 5746. .6163 of consensus"
repeat_region   54168. .54264
                /note="MIR repeat: matches 2. .111 of consensus"
misc_feature    complement(54484. .55028)
                /note="match: GSS: Em:AQ318586"
repeat_region   55956   56266
repeat_region   56471. .56682
                /note="AluSx repeat: matches 1. .312 of consensus"
repeat_region   56744. .56947
                /note="MIR repeat: matches 6. .232 of consensus"
repeat_region   57428. .57463
                /note="L2 repeat: matches 2246. .2448 of consensus"
misc_feature    complement(57599. .58119)
                /note="18 copies 2 mer tt 100% conserved"
repeat_region   58770. .59081
                /note="match: GSS: Em:AQ346372"
repeat_region   59703. .60000
                /note="AluY repeat: matches 1. .309 of consensus"
repeat_region   60141. .60194
                /note="AluY repeat: matches 1. .294 of consensus"
repeat_region   61142. .61188
                /note="L2 repeat: matches 2652. .2706 of consensus"
repeat_region   61806. .61910
                /note="L2 repeat: matches 2652. .2698 of consensus"

Query Match             4.3%;  Score 26;  DB 93;  Length 105137;
Best Local Similarity   100.0%;  Pred. No. 0.0013;
Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 tgctactggcatctaatgcatagagg 110
           ||||||||||||||||||||||||||
Db   66659 TGCTACTGGCATCTAATGCATAGAGG 66684


RESULT   4
AC012443
LOCUS       AC012443   154969 bp    DNA                HTG       03-APR-2001
DEFINITION  Homo sapiens chromosome UNK clone RP11-17G11, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC012443
VERSION     AC012443.7  GI:13518224
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154969)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 154969)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Apr 3, 2001 this sequence version replaced gi:13446343.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0017G11
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 34%
            Sequencing vector: plasmid; 53%
            Chemistry: Dye-primer ET; 34% of reads
            Chemistry: Dye-terminator Big Dye; 53% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 153554 bases at least Q40
            Consensus quality: 153754 bases at least Q30
            Consensus quality: 153862 bases at least Q20
            Insert size: 157000; agarose-fp
            Insert size: 153970; sum-of-contigs
            Quality coverage: 7.72 in Q20 bases; agarose-fp
            Quality coverage: 8.02 in Q20 bases; sum-of-contigs
            ----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
```

```
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   153970: contig of 153970 bp in length
            *  153971   154070: gap of unknown length
            *  154071   154969: contig of 899 bp in length.
FEATURES             Location/Qualifiers
     source          1. .154969
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="UNK"
                     /clone="RP11-17G11"
     misc_feature    1. .153970
                     /note="assembly_name:Contig23
                     clone_end:SP6
                     vector_side:right"
     misc_feature    154071. .154969
                     /note="assembly_name:Contig5"
BASE COUNT    48393 a  31230 c  30041 g  44839 t    466 others
ORIGIN


  Query Match             4.1%;  Score 25;  DB 62;  Length 154969;
  Best Local Similarity   100.0%;  Pred. No. 0.0051;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      75 gtggggaacatgctactggcatcta 99
           |||||||||||||||||||||||||
Db   13126 GTGGGGAACATGCTACTGGCATCTA 13150


RESULT   5
AC090051
LOCUS       AC090051   183893 bp    DNA              HTG       04-MAR-2001
DEFINITION  Homo sapiens chromosome 12q clone RP11-467E3, WORKING DRAFT
            SEQUENCE, 28 unordered pieces.
ACCESSION   AC090051
VERSION     AC090051.2  GI:13194926
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 183893)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G , Metzker,M , Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
            and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 183893)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 4, 2001 this sequence version replaced gi:12746498.
            ------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ------------- Project Information
                Center project name: HCYV
                Center clone name: RP11-467E3
            ------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 163729 bases at least Q40
                Consensus quality: 171892 bases at least Q30
                Consensus quality: 175386 bases at least Q20
                Estimated insert size: 174626; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 3.8x in Q20 bases; sum-of-contigs estimation
            -------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 28 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   25049: contig of 25049 bp in length
            *   25050   25149: gap of unknown length
            *   25150   45366: contig of 20217 bp in length
            *   45367   45466: gap of unknown length
            *   45467   65230: contig of 19764 bp in length
            *   65231   65330: gap of unknown length
            *   65331   75279: contig of 9949 bp in length
            *   75280   75379: gap of unknown length
            *   75380   85319: contig of 9940 bp in length
            *   85320   85419: gap of unknown length
            *   85420   95758: contig of 10339 bp in length
            *   95759   95858: gap of unknown length
            *   95859  105252: contig of 9394 bp in length
            *  105253  105352: gap of unknown length
            *  105353  115131: contig of 9779 bp in length
            *  115132  115231: gap of unknown length
            *  115232  123166: contig of 7935 bp in length
            *  123167  123266: gap of unknown length
            *  123267  130454: contig of 7188 bp in length
            *  130455  130554: gap of unknown length
            *  130555  134611: contig of 4057 bp in length
```

```
          *   134612   134711: gap of unknown length
          *   134712   138208: contig of 3497 bp in length
          *   138209   138308: gap of unknown length
          *   138309   142509: contig of 4201 bp in length
          *   142510   142609: gap of unknown length
          *   142610   145816: contig of 3207 bp in length
          *   145817   145916: gap of unknown length
          *   145917   148573: contig of 2657 bp in length
          *   148574   148673: gap of unknown length
          *   148674   152584: contig of 3911 bp in length
          *   152585   152684: gap of unknown length
          *   152685   156305: contig of 3621 bp in length
          *   156306   156405: gap of unknown length
          *   156406   159712: contig of 3307 bp in length
          *   159713   159812: gap of unknown length
          *   159813   161776: contig of 1964 bp in length
          *   161777   161876: gap of unknown length
          *   161877   165060: contig of 3184 bp in length
          *   165061   165160: gap of unknown length
          *   165161   168260: contig of 3100 bp in length
          *   168261   168360: gap of unknown length
          *   168361   171531: contig of 3171 bp in length
          *   171532   171631: gap of unknown length
          *   171632   174025: contig of 2394 bp in length
          *   174026   174125: gap of unknown length
          *   174126   176898: contig of 2773 bp in length
          *   176899   176998: gap of unknown length
          *   176999   179129: contig of 2131 bp in length
          *   179130   179229: gap of unknown length
          *   179230   181601: contig of 2372 bp in length
          *   181602   181701: gap of unknown length
          *   181702   182782: contig of 1081 bp in length
          *   182783   182882: gap of unknown length
          *   182883   183893: contig of 1011 bp in length.
FEATURES             Location/Qualifiers
     source          1. .183893
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12q"
                     /clone="RP11-467E3"
BASE COUNT    54634 a  36554 c  36547 g  52928 t   3230 others
ORIGIN


  Query Match             4.1%;  Score 25;  DB 77;  Length 183893;
  Best Local Similarity   100.0%;  Pred. No. 0.0051;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      86 gctactggcatctaatgcatagagg 110
           |||||||||||||||||||||||||
Db   37859 GCTACTGGCATCTAATGCATAGAGG 37883


RESULT   6
AC079344
LOCUS       AC079344   194294 bp    DNA              HTG       31-AUG-2000
DEFINITION  Homo sapiens chromosome UNK clone RP11-150L12, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC079344
VERSION     AC079344.2  GI:9954828
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 194294)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 194294)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 31, 2000 this sequence version replaced gi:9931950.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0150L12
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 184786 bases at least Q40
            Consensus quality: 188183 bases at least Q30
            Consensus quality: 189812 bases at least Q20
            Insert size: 188000; agarose-fp
            Insert size: 193094; sum-of-contigs
            Quality coverage: 5.74 in Q20 bases; agarose-fp
            Quality coverage: 5.64 in Q20 bases; sum-of-contigs
            ----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1277: contig of 1277 bp in length
            *     1278     1377: gap of unknown length
            *     1378    11800: contig of 10423 bp in length
            *    11801    11900: gap of unknown length
            *    11901    13353: contig of 1453 bp in length
            *    13354    13453: gap of unknown length
            *    13454    19171: contig of 5718 bp in length
            *    19172    19271: gap of unknown length
            *    19272    32004: contig of 12733 bp in length
            *    32005    32104: gap of unknown length
            *    32105    43844: contig of 11740 bp in length
            *    43845    43944: gap of unknown length
            *    43945    53017: contig of 9073 bp in length
            *    53018    53117: gap of unknown length
            *    53118    64048: contig of 10931 bp in length
            *    64049    64148: gap of unknown length
            *    64149    77841: contig of 13693 bp in length
            *    77842    77941: gap of unknown length
            *    77942    95865: contig of 17924 bp in length
            *    95866    95965: gap of unknown length
            *    95966   122392: contig of 26427 bp in length
            *   122393   122492: gap of unknown length
            *   122493   153546: contig of 31054 bp in length
            *   153547   153646: gap of unknown length
            *   153647   194294: contig of 40648 bp in length.
FEATURES             Location/Qualifiers
     source          1. .194294
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="UNK"
                     /clone="RP11-150L12"
     misc_feature    1. .1277
                     /note="assembly_name:Contig11
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1378. .11800
                     /note="assembly_name:Contig15
                     clone_end:T7
                     vector_side:right"
     misc_feature    11901. .13353
                     /note="assembly_name:Contig12"
```

```
     misc_feature    13454. .19171
                     /note="assembly_name:Contig13"
     misc_feature    19272. .32004
                     /note="assembly_name:Contig14"
     misc_feature    32105. .43844
                     /note="assembly_name:Contig16"
     misc_feature    43945. .53017
                     /note="assembly_name:Contig17"
     misc_feature    53118. .64048
                     /note="assembly_name:Contig18"
     misc_feature    64149. .77841
                     /note="assembly_name:Contig19"
     misc_feature    77942. .95865
                     /note="assembly_name:Contig20"
     misc_feature    95966. .122392
                     /note="assembly_name:Contig21"
     misc_feature    122493. .153546
                     /note="assembly_name:Contig22"
     misc_feature    153647. .194294
                     /note="assembly_name:Contig23"
BASE COUNT    60393 a  37283 c  37173 g  58245 t   1200 others
ORIGIN

  Query Match             4.1%;  Score 25;  DB 76;  Length 194294;
  Best Local Similarity   100.0%;  Pred. No. 0.0052;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      75 gtggggaacatgctactggcatcta 99
           |||||||||||||||||||||||||
Db   73447 GTGGGGAACATGCTACTGGCATCTA 73471


RESULT   7
AC090599/c
LOCUS       AC090599     67084 bp    DNA              HTG       03-MAR-2001
DEFINITION  Homo sapiens chromosome 11 clone RP11-674E24 map 11, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC090599
VERSION     AC090599.1  GI:13194332
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 67084)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-674E24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 67084)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
               Center project name: L11854
               Center clone name: 674_E_24
            --------------.
            * NOTE: This record contains 82 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      732: contig of 732 bp in length
            *      733  832: gap of      100 bp
            *      833     1536: contig of 704 bp in length
            *     1537  1636: gap of      100 bp
            *     1637     2355: contig of 719 bp in length
            *     2356  2455: gap of      100 bp
            *     2456     3173: contig of 718 bp in length
            *     3174  3273: gap of      100 bp
            *     3274     3969: contig of 696 bp in length
            *     3970  4069: gap of      100 bp
            *     4070     4793: contig of 724 bp in length
            *     4794  4893: gap of      100 bp
            *     4894     5626: contig of 733 bp in length
            *     5627  5726: gap of      100 bp
            *     5727     6445: contig of 719 bp in length
            *     6446  6545: gap of      100 bp
            *     6546     7262: contig of 717 bp in length
            *     7263  7362: gap of      100 bp
            *     7363     8088: contig of 726 bp in length
            *     8089  8188: gap of      100 bp
            *     8189     8905: contig of 717 bp in length
            *     8906  9005: gap of      100 bp
            *     9006     9725: contig of 720 bp in length
            *     9726  9825: gap of      100 bp
            *     9826    10539: contig of 714 bp in length
            *    10540 10639: gap of      100 bp
            *    10640    11354: contig of 715 bp in length
            *    11355 11454: gap of      100 bp
            *    11455    12164: contig of 710 bp in length
            *    12165 12264: gap of      100 bp
            *    12265    12974: contig of 710 bp in length
            *    12975 13074: gap of      100 bp
            *    13075    13806: contig of 732 bp in length
            *    13807 13906: gap of      100 bp
            *    13907    14628: contig of 722 bp in length
            *    14629 14728: gap of      100 bp
            *    14729    15460: contig of 732 bp in length
            *    15461 15560: gap of      100 bp
            *    15561    16277: contig of 717 bp in length
            *    16278 16377: gap of      100 bp
            *    16378    17115: contig of 738 bp in length
            *    17116 17215: gap of      100 bp
            *    17216    17922: contig of 707 bp in length
            *    17923 18022: gap of      100 bp
            *    18023    18732: contig of 710 bp in length
            *    18733 18832: gap of      100 bp
            *    18833    19546: contig of 714 bp in length
            *    19547 19646: gap of      100 bp
            *    19647    20358: contig of 712 bp in length
```

```
*    20359 20458: gap of      100 bp
*    20459    21188: contig of 730 bp in length
*    21189 21288: gap of      100 bp
*    21289    22004: contig of 716 bp in length
*    22005 22104: gap of      100 bp
*    22105    22805: contig of 701 bp in length
*    22806 22905: gap of      100 bp
*    22906    23633: contig of 728 bp in length
*    23634 23733: gap of      100 bp
*    23734    24453: contig of 720 bp in length
*    24454 24553: gap of      100 bp
*    24554    25281: contig of 728 bp in length
*    25282 25381: gap of      100 bp
*    25382    26107: contig of 726 bp in length
*    26108 26207: gap of      100 bp
*    26208    26925: contig of 718 bp in length
*    26926 27025: gap of      100 bp
*    27026    27748: contig of 723 bp in length
*    27749 27848: gap of      100 bp
*    27849    28576: contig of 728 bp in length
*    28577 28676: gap of      100 bp
*    28677    29417: contig of 741 bp in length
*    29418 29517: gap of      100 bp
*    29518    30229: contig of 712 bp in length
*    30230 30329: gap of      100 bp
*    30330    31033: contig of 704 bp in length
*    31034 31133: gap of      100 bp
*    31134    31844: contig of 711 bp in length
*    31845 31944: gap of      100 bp
*    31945    32640: contig of 696 bp in length
*    32641 32740: gap of      100 bp
*    32741    33442: contig of 702 bp in length
*    33443 33542: gap of      100 bp
*    33543    34261: contig of 719 bp in length
*    34262 34361: gap of      100 bp
*    34362    35079: contig of 718 bp in length
*    35080 35179: gap of      100 bp
*    35180    35913: contig of 734 bp in length
*    35914 36013: gap of      100 bp
*    36014    36743: contig of 730 bp in length
*    36744 36843: gap of      100 bp
*    36844    37575: contig of 732 bp in length
*    37576 37675: gap of      100 bp
*    37676    38393: contig of 718 bp in length
*    38394 38493: gap of      100 bp
*    38494    39213: contig of 720 bp in length
*    39214 39313: gap of      100 bp
*    39314    40023: contig of 710 bp in length
*    40024 40123: gap of      100 bp
*    40124    40846: contig of 723 bp in length
*    40847 40946: gap of      100 bp
*    40947    41641: contig of 695 bp in length
*    41642 41741: gap of      100 bp
*    41742    42459: contig of 718 bp in length
*    42460 42559: gap of      100 bp
*    42560    43258: contig of 699 bp in length
*    43259 43358: gap of      100 bp
*    43359    44090: contig of 732 bp in length
*    44091 44190: gap of      100 bp
*    44191    44920: contig of 730 bp in length
*    44921 45020: gap of      100 bp
*    45021    45739: contig of 719 bp in length
*    45740 45839: gap of      100 bp
*    45840    46561: contig of 722 bp in length
*    46562 46661: gap of      100 bp
*    46662    47378: contig of 717 bp in length
*    47379 47478: gap of      100 bp
*    47479    48194: contig of 716 bp in length
*    48195 48294: gap of      100 bp
*    48295    49024: contig of 730 bp in length
*    49025 49124: gap of      100 bp
*    49125    49845: contig of 721 bp in length
*    49846 49945: gap of      100 bp
*    49946    50658: contig of 713 bp in length
*    50659 50758: gap of      100 bp
*    50759    51472: contig of 714 bp in length
*    51473 51572: gap of      100 bp
*    51573    52284: contig of 712 bp in length
*    52285 52384: gap of      100 bp
*    52385    53101: contig of 717 bp in length
*    53102 53201: gap of      100 bp
*    53202    53935: contig of 734 bp in length
*    53936 54035: gap of      100 bp
*    54036    54767: contig of 732 bp in length
*    54768 54867: gap of      100 bp
*    54868    55587: contig of 720 bp in length
*    55588 55687: gap of      100 bp
*    55688    56419: contig of 732 bp in length
*    56420 56519: gap of      100 bp

Query Match              3.8%;  Score 23;  DB 78;  Length 67084;
Best Local Similarity   100.0%;  Pred. No. 0.075;
Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     163 agaattatctagccccaaatgtc 185
           |||||||||||||||||||||||
Db   33966 AGAATTATCTAGCCCCAAATGTC 33944


RESULT   8
HS106C24
LOCUS       HS106C24    93959 bp    DNA                 PRI       23-NOV-1999
DEFINITION  Human DNA sequence from PAC 106C24, between markers DXS294 and
            DXS730 on chromosome X.
ACCESSION   Z83313
VERSION     Z83313.1  GI:1730462
KEYWORDS    X.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 93959)
  AUTHORS   Isherwood,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-DEC-1996) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquires: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     IMPORTANT:  This sequence is not the entire insert of clone 106C24.
             It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu).  Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true left end of clone 106C24 is at 1 in this sequence. The
            true right end of clone 206J9 is at 47751.
            106C24 is from a whole genome PAC library.
FEATURES             Location/Qualifiers
     source          1. .93959
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="X"
                     /clone="RP1-106C24"
                     /clone_lib="RPCI-1"
     repeat_region   552. .799
                     /note="MIR repeat: matches 259. .3 of consensus"
     repeat_region   2498. .2582
                     /note="MIR repeat: matches 57. .145 of consensus"
     repeat_region   3438. .3504
                     /note="MIR repeat: matches 122. .56 of consensus"
     repeat_region   4744. .4938
                     /note="MER3 repeat: matches 2. .200 of consensus"
```

repeat_region 5365. .5599
/note="MIR repeat: matches 8. .236 of consensus"
repeat_region 7810. .7942
/note="MIR repeat: matches 13. .154 of consensus"
repeat_region 8410. .8520
/note="MIR repeat: matches 11. .119 of consensus"
repeat_region 9006. .9089
/note="MIR2 repeat: matches 63. .146 of consensus"
repeat_region 9575. .9767
/note="L1 repeat: matches 5020. .4829 of consensus"
repeat_region 10196. .10275
/note="MER5A repeat: matches 106. .4 of consensus"
repeat_region 10515. .10619
/note="MIR repeat: matches 49. .153 of consensus"
repeat_region 11164. .11350
/note="L1ME3A repeat: matches 702. .907 of consensus"
repeat_region 12097. .12302
/note="MIR repeat: matches 30. .262 of consensus"
repeat_region 12938. .13023
/note="MIR2 repeat: matches 146. .64 of consensus"
repeat_region 14787. .14844
/note="29 copies of 2 mer 90 % conserved"
repeat_region 17057. .17086
/note="10 copies of 3 mer 93 % conserved"
repeat_region 17105. .17343
/note="MIR repeat: matches 258. .11 of consensus"
repeat_region 17750. .18041
/note="AluSx repeat: matches 297. .5 of consensus"
repeat_region 19193. .19360
/note="MIR repeat: matches 213. .49 of consensus"
repeat_region 20826. .20962
/note="L1 repeat: matches 1. .136 of consensus"
repeat_region 20964. .21104
/note="L1MA1 repeat: matches 92. .234 of consensus"
repeat_region 21093. .21869
/note="L1 repeat: matches 5. .772 of consensus"
repeat_region 21856. .26363
/note="L1 repeat: matches 887. .5390 of consensus"
repeat_region 26214. .27104
/note="L1PA2 repeat: matches 1. .891 of consensus"
repeat_region 29529. .29582
/note="MIR repeat: matches 66. .120 of consensus"
repeat_region 30103. .30321
/note="MLT1A2 repeat: matches 374. .156 of consensus"
repeat_region 30339. .30589
/note="L1PA8 repeat: matches 909. .659 of consensus"
repeat_region 30596. .30743
/note="MLT1A1 repeat: matches 149. .1 of consensus"
repeat_region 30834. .31103
/note="MLT1-INTERNAL repeat: matches 1231. .961 ofconsensus"
repeat_region 31580. .31951
/note="MLT1-INTERNAL repeat: matches 599. .231 of consensus"
repeat_region 31957. .32266
/note="AluSc repeat: matches 297. .1 of consensus"
repeat_region 32267. .32482
/note="MLT1-INTERNAL repeat: matches 249. .5 of consensus"
repeat_region 32487. .32545
/note="MLT1A2 repeat: matches 374. .313 of consensus"
repeat_region 32551. .32850
/note="AluY repeat: matches 1. .301 of consensus"
repeat_region 32870. .32911
/note="MSTD repeat: matches 270. .312 of consensus"
repeat_region 32912. .33202
/note="MLT1A1 repeat: matches 297. .1 of consensus"
repeat_region 33703. .33777
/note="MER2 repeat: matches 345. .271 of consensus"
repeat_region 33811. .33871
/note="L1MA3 repeat: matches 990. .1050 of consensus"
repeat_region 33898. .34115
/note="MER2 repeat: matches 210. .1 of consensus"
repeat_region 35085. .35213
/note="MIR repeat: matches 152. .29 of consensus"
repeat_region 35255. .35321
/note="MIR repeat: matches 149. .84 of consensus"
repeat_region 35621. .36097
/note="MLT1C repeat: matches 1. .465 of consensus"
repeat_region 36588. .36721
/note="MIR repeat: matches 119. .256 of consensus"
repeat_region 38850. .39742
/note="L1PA2 repeat: matches 893. .1 of consensus"
repeat_region 39598. .44765
/note="L1 repeat: matches 5390. .386 of consensus"
repeat_region 45104. .45334
/note="L1 repeat: matches 255. .12 of consensus"
repeat_region 48462. .48649
/note="L1MA10 repeat: matches 935. .737 of consensus"
repeat_region 50125. .53643
/note="L1 repeat: matches 1877. .5390 of consensus"
repeat_region 53494. .54383
/note="L1PA2 repeat: matches 1. .891 of consensus"
repeat_region 54743. .55195
/note="L1MB7 repeat: matches 867. .393 of consensus"
repeat_region 55530. .56049
/note="L1PA10 repeat: matches 889. .311 of consensus"
repeat_region 57142. .57399
/note="L1 repeat: matches 5129. .5389 of consensus"
repeat_region 57251. .58278
/note="L1MA9 repeat: matches 1. .1055 of consensus"
repeat_region 58483. .58557
/note="L1PA2 repeat: matches 815. .889 of consensus"
repeat_region 58564. .58597
/note="17 copies of 2 mer 82 % conserved"
repeat_region 58905. .59344
/note="L1MC2 repeat: matches 12. .484 of consensus"
repeat_region 59359. .59409
/note="MER42c repeat: matches 1488. .1538 of consensus"
repeat_region 59440. .59664
/note="MER30 repeat: matches 230. .1 of consensus"
repeat_region 60853. .60918
/note="MIR2 repeat: matches 78. .143 of consensus"
repeat_region 61282. .61335
/note="27 copies of 2 mer 96 % conserved"
repeat_region 61374. .61454
/note="MIR repeat: matches 146. .64 of consensus"
repeat_region 61902. .62103
/note="MIR repeat: matches 9. .247 of consensus"
repeat_region 64007. .64146
/note="MER3 repeat: matches 32. .170 of consensus"
repeat_region 64287. .64532
/note="MER33 repeat: matches 70. .324 of consensus"
repeat_region 67972. .68217
/note="MIR repeat: matches 4. .247 of consensus"
repeat_region 68347. .68471
/note="MIR repeat: matches 15. .144 of consensus"
repeat_region 70707. .70748
/note="MIR repeat: matches 88. .47 of consensus"
repeat_region 71075. .71375
/note="AluY repeat: matches 1. .300 of consensus"
repeat_region 71608. .71649
/note="MIR2 repeat: matches 141. .100 of consensus"
repeat_region 71736. .71944
/note="MSTD repeat: matches 5. .215 of consensus"
repeat_region 71985. .72402
/note="MSTB repeat: matches 2. .426 of consensus"
repeat_region 72407. .72502
/note="MSTD repeat: matches 233. .328 of consensus"
repeat_region 72503. .72804
/note="AluSx repeat: matches 301. .2 of consensus"
repeat_region 72806. .72876
/note="MSTD repeat: matches 324. .394 of consensus"
repeat_region 72877. .72979
/note="MST-INTERNAL repeat: matches 1. .102 of consensus"
repeat_region 72909. .73456
/note="MLT1-INTERNAL repeat: matches 32. .593 of consensus"

```
     repeat_region   73468. .73517
                     /note="25 copies of 2 mer 96 % conserved"
     repeat_region   73543. .74432
                     /note="MST-INTERNAL repeat: matches 532. .1458 of
                     consensus"
     repeat_region   74466. .75346
                     /note="L1PA12 repeat: matches 912. .4 of consensus"
     repeat_region   75199. .79854
                     /note="L1 repeat: matches 5390. .767 of consensus"

  Query Match              3.8%;  Score 23;  DB 92;  Length 93959;
  Best Local Similarity   100.0%;  Pred. No. 0.076;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     163 agaattatctagccccaaatgtc 185
           |||||||||||||||||||||||
Db   85580 AGAATTATCTAGCCCCAAATGTC 85602


RESULT  9
AC073940/c
LOCUS       AC073940   103757 bp    DNA                HTG       14-APR-2001
DEFINITION  Homo sapiens chromosome 15 clone CTD-2190J1 map 15q15, ***
            SEQUENCING IN PROGRESS ***, 1 ordered pieces.
ACCESSION   AC073940
VERSION     AC073940.3  GI:13624386
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 103757)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE     Sequencing of human chromosome 15 D15S146-D15S117 region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 103757)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A.,
            Madan,A., Nesbitt,R., Shaffer,T. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-2000) Multimegabase Sequencing Center, Institute
            for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
            98105, USA
COMMENT     On Apr 14, 2001 this sequence version replaced gi:12658005.
            ------------ Genome Center
                Center: Multimegabase Sequencing Center
                Center code: UWMSC
                Web site: http://chroma.mbt.washington.edu/msg_www
                Contact: leerowen@systemsbiology.org
            ------------ Summary Statistics
                Sequencing vector: pUC18; L08752
                Chemistry: Dye-terminator Big Dye; 90% of reads
                Chemistry: Dye-primer Big Dye; 10% of reads
                Assembly program: Phrap; version 0.990399
                Insert size: 100000; agarose-fp
                Quality coverage: 10.2x in Q20 bases; sum-of-contigs
            ------------.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            ---------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1   103757: contig of 103757 bp in length.
FEATURES              Location/Qualifiers
     source           1. .103757
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="15"
                      /map="15q15"
                      /clone="CTD-2190J1"
                      /clone_lib="Cal Tech human BAC library D"
                      /note="This clone overlaps RP11-213B23 and RP11-616K22"
BASE COUNT    32585 a  21521 c  21252 g  28399 t
ORIGIN

  Query Match              3.8%;  Score 23;  DB 75;  Length 103757;
  Best Local Similarity   100.0%;  Pred. No. 0.077;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     163 agaattatctagccccaaatgtc 185
           |||||||||||||||||||||||
Db   94438 AGAATTATCTAGCCCCAAATGTC 94416


RESULT  10
AC002554
LOCUS       AC002554   127590 bp    DNA                PRI       27-AUG-1998
DEFINITION  Human Chromosome 11p11.2 PAC clone pDJ404m15, complete sequence.
ACCESSION   AC002554
VERSION     AC002554.1  GI:3478647
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 127590)
  AUTHORS   Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,
            Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J.,
            English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
            Grant,O., Hahner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N.,
            Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,
            Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
  TITLE     HTGS Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 127590)
  AUTHORS   Evans,G.A., Athanasiou,M., Bradbury,P., Brignac,S., Bumeister,R.,
            Davis,C., English,C., Franklin,T.L., Garner,H.R., Gordon,M.,
            Gotway,G., Grant,O., Hahner,L., Harris,J., Hinson,S.,
            Narayanaswamy,U., Newton,J., O'Brien,K., Oliver,T., Patel,P.,
            Probst,S., Rayner,S., Schageman,J., Schilling,P., Schultz,R.,
            Syed,M., Valenzuela,D., Ward,T. and Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-1997) Genome Science and Technology Center,
            University of Texas Southwestern Medical Center at Dallas, 5323
            Harry Hines Blvd, Dallas, TX 75235-8591, USA
REFERENCE   3  (bases 1 to 127590)
  AUTHORS   Evans,G.A., Athanasiou,M., Basit,M., Bradbury,P., Brignac,S.,
            Bumesiter,R., Davis,C., English,C., Franklin,T.L., Garner,H.R.,
            Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J.,
            Hinson,S., Narayanaswamy,U., Newton,J., O'Brien,K., Patel,P.,
            Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D.,
            Ward,T. and Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1997) Genome Science & Technology Center,
            University of Texas Southwester Medical Center, 5323 Harry Hines
            Blvd, Dallas, TX 75235-8591, USA
REFERENCE   4  (bases 1 to 127590)
  AUTHORS   Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,
            Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J.,
```

```
            English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
            Grant,O., Hahner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N.,
            Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,
            Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1998) Genome Science & Technology Center,
            University of Texas Southwestern Medical Center, 5323 Harry Hines
            Blvd, Dallas, TX 75235-8591, USA
COMMENT     On Aug 27, 1998 this sequence version replaced gi:2695561.
            CHROMSOMAL LOCUS: This PAC clone comes from the 11p11.2
            region mapped between STSs markers D11S578 and D11S3805h.
            MARKER CONFIRMATION: STS sequence confirmed; D11S578
            MAPPED CLONE OVERLAP: pDJ404c10.
            IMPORTANT: This submission contains the entire insert of clone
            pDJ404m15.  pDJ404m15 comes from the RPCI-3 PAC library constructed
            at the Roswell Park Cancer Institute by the Pieter de Jong group.
            This clone has been finished according to strict quality criteria
            and attempts have been made to resolve all base calling problems
            such as compressions and repetitive elements. The expected
            Phred/Phrap calculated errors/10kb is 0.23.  In addition, this
            sequence has been finished such that 99.9% of consensus base calls
            consist of either double-stranded coverage or 2 types of labeling
            chemistry on one strand. Additional information regarding this
            clone may be found on the GESTEC web page at
            http://gestec.swmed.edu/11p11.htm.
FEATURES            Location/Qualifiers
     source         1. .127590
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
     repeat_region  complement(1. .69)
                    /rpt_family="MER41"
     repeat_region  90. .385
                    /rpt_family="Alu"
     repeat_region  complement(400. .821)
                    /rpt_family="MER41"
     repeat_region  complement(1101. .1382)
                    /rpt_family="LTR9"
     repeat_region  complement(1436. .1731)
                    /rpt_family="Alu"
     repeat_region  complement(1797. .1937)
                    /rpt_family="LTR9"
     repeat_region  complement(3274. .3558)
                    /rpt_family="Alu"
     repeat_region  complement(3923. .4100)
                    /rpt_family="MER4"
     repeat_region  4779. .5469
                    /rpt_family="LTR8"
     repeat_region  complement(7262. .7513)
                    /rpt_family="Alu"
     repeat_region  complement(7651. .7964)
                    /rpt_family="Alu"
     repeat_region  11021. .11306
                    /rpt_family="Alu"
     repeat_region  11424. .12625
                    /rpt_family="L1"
     repeat_region  13008. .13297
                    /rpt_family="Alu"
     repeat_region  13356. .13506
                    /rpt_family="L1"
     repeat_region  complement(14032. .14334)
                    /rpt_family="Alu"
     repeat_region  14861. .15068
                    /rpt_family="MER20"
     repeat_region  15934. .16238
                    /rpt_family="Alu"
     repeat_region  16914. .17018
                    /rpt_family="Alu"
     repeat_region  complement(17093. .17337)
                    /rpt_family="Alu"
     repeat_region  complement(17362. .17691)
                    /rpt_family="MER42"
     repeat_region  17918. .18156
                    /rpt_family="Alu"
     repeat_region  complement(18206. .18402)
                    /rpt_family="MIR"
     repeat_region  19740. .19807
                    /rpt_family="Alu"
     repeat_region  complement(21285. .21395)
                    /rpt_family="MLT1"
     repeat_region  21587. .21762
                    /rpt_family="MLT1"
     repeat_region  21779. .21846
                    /rpt_family="MLT1"
     repeat_region  21857. .21924
                    /rpt_family="MSTAR"
     repeat_region  23972. .24546
                    /rpt_family="MLT1"
     repeat_region  24638. .24733
                    /rpt_family="MIR"
     repeat_region  complement(24860. .25148)
                    /rpt_family="Alu"
     repeat_region  26031. .26203
                    /rpt_family="Alu"
     repeat_region  26528. .26717
                    /rpt_family="Alu"
     repeat_region  complement(27650. .27708)
                    /rpt_family="MIR"
     repeat_region  complement(29082. .29343)
                    /rpt_family="Alu"
     repeat_region  30684. .30980
                    /rpt_family="Alu"
     repeat_region  32634. .32818
                    /rpt_family="Alu"
     repeat_region  complement(35666. .35779)
                    /rpt_family="Alu"
     repeat_region  36025. .36096
                    /rpt_family="MIR"
     repeat_region  36923. .37030
                    /rpt_family="MIR"
     repeat_region  complement(37101. .37189)
                    /rpt_family="THE1"
     repeat_region  37653. .37940
                    /rpt_family="Alu"
     repeat_region  38374. .38647
                    /rpt_family="Alu"
     repeat_region  complement(38907. .39198)
                    /rpt_family="Alu"
     repeat_region  complement(40840. .40945)
                    /rpt_family="MER5"
     repeat_region  complement(41499. .41785)
                    /rpt_family="Alu"
     repeat_region  42735. .43073
                    /rpt_family="MLT1"
     repeat_region  complement(43666. .43951)
                    /rpt_family="Alu"
     repeat_region  44899. .45060
                    /rpt_family="MIR"
     repeat_region  complement(45617. .45784)
                    /rpt_family="Alu"
     repeat_region  45878. .46030
                    /rpt_family="MIR"
     repeat_region  complement(46105. .46157)
                    /rpt_family="MLT1"
     repeat_region  complement(46177. .46305)
                    /rpt_family="MLT1"
     repeat_region  46676. .46841
                    /rpt_family="MER2"
     repeat_region  46871. .46959
                    /rpt_family="Alu"
     repeat_region  complement(46983. .47265)
                    /rpt_family="Alu"
     repeat_region  47306. .47395
                    /rpt_family="Alu"
     repeat_region  47503. .47580
                    /rpt_family="Alu"
     repeat_region  47627. .47741
```

```
                /rpt_family="MER2"
     repeat_region   complement(47741. .47839)
                /rpt_family="MIR"
     repeat_region   49558. .49671
                /rpt_family="MIR"
     repeat_region   50232. .50508
                /rpt_family="Alu"
     repeat_region   complement(52633. .52764)
                /rpt_family="MIR"
     repeat_region   complement(53356. .53634)
                /rpt_family="Alu"

  Query Match              3.8%;  Score 23;  DB 85;  Length 127590;
  Best Local Similarity   100.0%;  Pred. No. 0.078;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    163 agaattatctagccccaaatgtc 185
          |||||||||||||||||||||||
Db  15032 AGAATTATCTAGCCCCAAATGTC 15054


RESULT  11
AC067916/c
LOCUS       AC067916   134394 bp    DNA             HTG       09-JUN-2000
DEFINITION  Homo sapiens chromosome X clone RP11-513J24 map X, *** SEQUENCING
            IN PROGRESS ***, 32 unordered pieces.
ACCESSION   AC067916
VERSION     AC067916.2  GI:8389561
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 134394)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome X, clone RP11-513J24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 134394)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
  COMMENT   On Jun 9, 2000 this sequence version replaced gi:7652204.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6355
                Center clone name: 513_J_24
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 32 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1     1240: contig of 1240 bp in length
            *   1241  1340: gap of      100 bp
            *   1341     2613: contig of 1273 bp in length
            *   2614  2713: gap of      100 bp
            *   2714     4159: contig of 1446 bp in length
            *   4160  4259: gap of      100 bp
            *   4260     5435: contig of 1176 bp in length
            *   5436  5535: gap of      100 bp
            *   5536     6901: contig of 1366 bp in length
            *   6902  7001: gap of      100 bp
            *   7002     8119: contig of 1118 bp in length
            *   8120  8219: gap of      100 bp
            *   8220    10213: contig of 1994 bp in length
            *  10214 10313: gap of      100 bp
            *  10314    12093: contig of 1780 bp in length
            *  12094 12193: gap of      100 bp
            *  12194    13509: contig of 1316 bp in length
            *  13510 13609: gap of      100 bp
            *  13610    14871: contig of 1262 bp in length
            *  14872 14971: gap of      100 bp
            *  14972    16966: contig of 1995 bp in length
            *  16967 17066: gap of      100 bp
            *  17067    18732: contig of 1666 bp in length
            *  18733 18832: gap of      100 bp
            *  18833    20655: contig of 1823 bp in length
            *  20656 20755: gap of      100 bp
            *  20756    22721: contig of 1966 bp in length
            *  22722 22821: gap of      100 bp
            *  22822    24991: contig of 2170 bp in length
            *  24992 25091: gap of      100 bp
            *  25092    25443: contig of 352 bp in length
            *  25444 25543: gap of      100 bp
            *  25544    28606: contig of 3063 bp in length
            *  28607 28706: gap of      100 bp
            *  28707    30875: contig of 2169 bp in length
            *  30876 30975: gap of      100 bp
            *  30976    34223: contig of 3248 bp in length
            *  34224 34323: gap of      100 bp
            *  34324    37911: contig of 3588 bp in length
            *  37912 38011: gap of      100 bp
            *  38012    42033: contig of 4022 bp in length
            *  42034 42133: gap of      100 bp
            *  42134    45055: contig of 2922 bp in length
            *  45056 45155: gap of      100 bp
            *  45156    50793: contig of 5638 bp in length
            *  50794 50893: gap of      100 bp
            *  50894    56638: contig of 5745 bp in length
            *  56639 56738: gap of      100 bp
            *  56739    62187: contig of 5449 bp in length
            *  62188 62287: gap of      100 bp
            *  62288    69769: contig of 7482 bp in length
            *  69770 69869: gap of      100 bp
            *  69870    77546: contig of 7677 bp in length
            *  77547 77646: gap of      100 bp
            *  77647    84723: contig of 7077 bp in length
            *  84724 84823: gap of      100 bp
            *  84824    92916: contig of 8093 bp in length
            *  92917 93016: gap of      100 bp
            *  93017   107613: contig of 14597 bp in length
            * 107614 107713: gap of      100 bp
            * 107714   119925: contig of 12212 bp in length
            * 119926 120025: gap of      100 bp
            * 120026   134394: contig of 14369 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .134394
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="X"
                     /clone="RP11-513J24"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1240
                     /note="assembly_fragment"
     misc_feature    1341. .2613
                     /note="assembly_fragment"
     misc_feature    2714. .4159
                     /note="assembly_fragment"
     misc_feature    4260. .5435
                     /note="assembly_fragment"
     misc_feature    5536. .6901
                     /note="assembly_fragment"
     misc_feature    7002. .8119
                     /note="assembly_fragment"
     misc_feature    8220. .10213
                     /note="assembly_fragment"
     misc_feature    10314. .12093
                     /note="assembly_fragment"
     misc_feature    12194. .13509
                     /note="assembly_fragment"
     misc_feature    13610. .14871
                     /note="assembly_fragment"
     misc_feature    14972. .16966
                     /note="assembly_fragment"
     misc_feature    17067. .18732
                     /note="assembly_fragment"
     misc_feature    18833. .20655
                     /note="assembly_fragment"
     misc_feature    20756. .22721
                     /note="assembly_fragment"
     misc_feature    22822. .24991
                     /note="assembly_fragment"
     misc_feature    25092. .25443
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    25544. .28606
                     /note="assembly_fragment"
     misc_feature    28707. .30875
                     /note="assembly_fragment"
     misc_feature    30976. .34223
                     /note="assembly_fragment"
     misc_feature    34324. .37911
                     /note="assembly_fragment"
     misc_feature    38012. .42033
                     /note="assembly_fragment"
     misc_feature    42134. .45055
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    45156. .50793
                     /note="assembly_fragment"
     misc_feature    50894. .56638
                     /note="assembly_fragment"
     misc_feature    56739. .62187
                     /note="assembly_fragment"
     misc_feature    62288. .69769
                     /note="assembly_fragment"
     misc_feature    69870. .77546
                     /note="assembly_fragment"
     misc_feature    77647. .84723
                     /note="assembly_fragment"
     misc_feature    84824. .92916
                     /note="assembly_fragment"
     misc_feature    93017. .107613
                     /note="assembly_fragment"
     misc_feature    107714   119925
                     /note="assembly_fragment"
     misc_feature    120026. .134394
                     /note="assembly_fragment"
BASE COUNT    39328 a   25781 c   26014 g   40164 t    3107 others

 Query Match             3.8%;  Score 23;  DB 73;  Length 134394;
 Best Local Similarity   100.0%;  Pred. No. 0.078;
 Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      163 agaattatctagccccaaatgtc 185
            |||||||||||||||||||||||
Db    77870 AGAATTATCTAGCCCCAAATGTC 77848


RESULT  12
AC011573
LOCUS       AC011573   151357 bp    DNA              HTG       12-MAR-2000
DEFINITION  Homo sapiens clone RP11-12G12, WORKING DRAFT SEQUENCE, 28 unordered
            pieces.
ACCESSION   AC011573
VERSION     AC011573.3  GI:7230070
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151357)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-12G12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151357)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 12, 2000 this sequence version replaced gi:6249718.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3265
                Center clone name: 12_G_12
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 101368 bases at least Q40
                Consensus quality: 127052 bases at least Q30
                Consensus quality: 139101 bases at least Q20
                Insert size: 147000; agarose-fp
                Insert size: 148657; sum-of-contigs
                Quality coverage: 3.5 in Q20 bases; agarose-fp
                Quality coverage: 3.5 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
```

```
            * consists of 28 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1505: contig of 1505 bp in length
            *    1506 1605: gap of      100 bp
            *    1606     2995: contig of 1390 bp in length
            *    2996 3095: gap of      100 bp
            *    3096     4668: contig of 1573 bp in length
            *    4669 4768: gap of      100 bp
            *    4769     6111: contig of 1343 bp in length
            *    6112 6211: gap of      100 bp
            *    6212     7559: contig of 1348 bp in length
            *    7560 7659: gap of      100 bp
            *    7660     8904: contig of 1245 bp in length
            *    8905 9004: gap of      100 bp
            *    9005    10441: contig of 1437 bp in length
            *   10442 10541: gap of      100 bp
            *   10542    12992: contig of 2451 bp in length
            *   12993 13092: gap of      100 bp
            *   13093    16001: contig of 2909 bp in length
            *   16002 16101: gap of      100 bp
            *   16102    18808: contig of 2707 bp in length
            *   18809 18908: gap of      100 bp
            *   18909    22389: contig of 3481 bp in length
            *   22390 22489: gap of      100 bp
            *   22490    25729: contig of 3240 bp in length
            *   25730 25829: gap of      100 bp
            *   25830    28948: contig of 3119 bp in length
            *   28949 29048: gap of      100 bp
            *   29049    32582: contig of 3534 bp in length
            *   32583 32682: gap of      100 bp
            *   32683    36104: contig of 3422 bp in length
            *   36105 36204: gap of      100 bp
            *   36205    39825: contig of 3621 bp in length
            *   39826 39925: gap of      100 bp
            *   39926    45184: contig of 5259 bp in length
            *   45185 45284: gap of      100 bp
            *   45285    50916: contig of 5632 bp in length
            *   50917 51016: gap of      100 bp
            *   51017    56385: contig of 5369 bp in length
            *   56386 56485: gap of      100 bp
            *   56486    63439: contig of 6954 bp in length
            *   63440 63539: gap of      100 bp
            *   63540    70246: contig of 6707 bp in length
            *   70247 70346: gap of      100 bp
            *   70347    76319: contig of 5973 bp in length
            *   76320 76419: gap of      100 bp
            *   76420    85131: contig of 8712 bp in length
            *   85132 85231: gap of      100 bp
            *   85232    91974: contig of 6743 bp in length
            *   91975 92074: gap of      100 bp
            *   92075   102742: contig of 10668 bp in length
            *  102743 102842: gap of      100 bp
            *  102843   110539: contig of 7697 bp in length
            *  110540 110639: gap of      100 bp
            *  110640   119486: contig of 8847 bp in length
            *  119487 119586: gap of      100 bp
            *  119587   151357: contig of 31771 bp in length.
FEATURES             Location/Qualifiers
     source          1. .151357
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-12G12"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1505
                     /note="assembly_fragment"
     misc_feature    1606. .2995
                     /note="assembly_fragment"
     misc_feature    3096   4668
                     /note="assembly_fragment"
     misc_feature    4769. .6111
                     /note="assembly_fragment"
     misc_feature    6212. .7559
                     /note="assembly_fragment"
     misc_feature    7660. .8904
                     /note="assembly_fragment"
     misc_feature    9005. .10441
                     /note="assembly_fragment"
     misc_feature    10542. .12992
                     /note="assembly_fragment"
     misc_feature    13093. .16001
                     /note="assembly_fragment"
     misc_feature    16102. .18808
                     /note="assembly_fragment"
     misc_feature    18909. .22389
                     /note="assembly_fragment"
     misc_feature    22490. .25729
                     /note="assembly_fragment"
     misc_feature    25830. .28948
                     /note="assembly_fragment"
     misc_feature    29049. .32582
                     /note="assembly_fragment"
     misc_feature    32683. .36104
                     /note="assembly_fragment"
     misc_feature    36205. .39825
                     /note="assembly_fragment"
     misc_feature    39926. .45184
                     /note="assembly_fragment"
     misc_feature    45285. .50916
                     /note="assembly_fragment"
     misc_feature    51017. .56385
                     /note="assembly_fragment"
     misc_feature    56486. .63439
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    63540. .70246
                     /note="assembly_fragment"
     misc_feature    70347. .76319
                     /note="assembly_fragment"
     misc_feature    76420. .85131
                     /note="assembly_fragment"
     misc_feature    85232. .91974
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    92075. .102742
                     /note="assembly_fragment"
     misc_feature    102843. .110539
                     /note="assembly_fragment"
     misc_feature    110640. .119486
                     /note="assembly_fragment"
     misc_feature    119587. .151357
                     /note="assembly_fragment"
BASE COUNT    40770 a  33565 c  33412 g  40768 t   2842 others
ORIGIN


  Query Match             3.8%;  Score 23;  DB 62;  Length 151357;
  Best Local Similarity   100.0%;  Pred. No. 0.078;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      56 ttttgattgtcacacctgggtgg 78
           |||||||||||||||||||||||
Db  56633 TTTTGATTGTCACACCTGGGTGG 56655


RESULT  13
AC034161
LOCUS       AC034161   157402 bp    DNA              HTG       01-SEP-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-397J5, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
```

```
ACCESSION   AC034161
VERSION     AC034161.3  GI:9958241
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 157402)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 157402)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 1, 2000 this sequence version replaced gi:7523965.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0397J05
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 149716 bases at least Q40
            Consensus quality: 152398 bases at least Q30
            Consensus quality: 153456 bases at least Q20
            Insert size: 191000; agarose-fp
            Insert size: 155702; sum-of-contigs
            Quality coverage: 3.43 in Q20 bases; agarose-fp
            Quality coverage: 4.25 in Q20 bases; sum-of-contigs
            ----------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     5077: contig of 5077 bp in length
            *    5078     5177: gap of unknown length
            *    5178    11207: contig of 6030 bp in length
            *   11208    11307: gap of unknown length
            *   11308    18089: contig of 6782 bp in length
            *   18090    18189: gap of unknown length
            *   18190    24820: contig of 6631 bp in length
            *   24821    24920: gap of unknown length
            *   24921    32091: contig of 7171 bp in length
            *   32092    32191: gap of unknown length
            *   32192    45992: contig of 13801 bp in length
            *   45993    46092: gap of unknown length
            *   46093    52957: contig of 6865 bp in length
            *   52958    53057: gap of unknown length
            *   53058    61360: contig of 8303 bp in length
            *   61361    61460: gap of unknown length
            *   61461    69613: contig of 8153 bp in length
            *   69614    69713: gap of unknown length
            *   69714    80408: contig of 10695 bp in length
            *   80409    80508: gap of unknown length
            *   80509    91848: contig of 11340 bp in length
            *   91849    91948: gap of unknown length
            *   91949   115447: contig of 23499 bp in length
            *  115448   115547: gap of unknown length
            *  115548   142318: contig of 26771 bp in length
            *  142319   142418: gap of unknown length
            *  142419   145665: contig of 3247 bp in length
            *  145666   145765: gap of unknown length
            *  145766   148715: contig of 2950 bp in length
            *  148716   148815: gap of unknown length
            *  148816   151082: contig of 2267 bp in length
            *  151083   151182: gap of unknown length
            *  151183   153839: contig of 2657 bp in length
            *  153840   153939: gap of unknown length
            *  153940   157402: contig of 3463 bp in length.
FEATURES             Location/Qualifiers
     source          1. .157402
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="RP11-397J5"
     misc_feature    1. .5077
                     /note="assembly_name:Contig10"
     misc_feature    5178. .11207
                     /note="assembly_name:Contig11"
     misc_feature    11308. .18089
                     /note="assembly_name:Contig12"
     misc_feature    18190. .24820
                     /note="assembly_name:Contig13"
     misc_feature    24921. .32091
                     /note="assembly_name:Contig14
                     clone_end:SP6
                     vector_side:right"
     misc_feature    32192. .45992
                     /note="assembly_name:Contig19
                     clone_end:T7
                     vector_side:right"
     misc_feature    46093. .52957
                     /note="assembly_name:Contig15"
     misc_feature    53058. .61360
                     /note="assembly_name:Contig16"
     misc_feature    61461. .69613
                     /note="assembly_name:Contig17"
     misc_feature    69714. .80408
                     /note="assembly_name:Contig18"
     misc_feature    80509. .91848
                     /note="assembly_name:Contig20"
     misc_feature    91949. .115447
                     /note="assembly_name:Contig21"
     misc_feature    115548. .142318
                     /note="assembly_name:Contig22"
     misc_feature    142419. .145665
                     /note="assembly_name:Contig5"
     misc_feature    145766. .148715
                     /note="assembly_name:Contig6"
     misc_feature    148816. .151082
                     /note="assembly_name:Contig7"
     misc_feature    151183. .153839
                     /note="assembly_name:Contig8"
     misc_feature    153940. .157402
                     /note="assembly_name:Contig9"
BASE COUNT    49815 a  29500 c  29459 g  46922 t   1706 others
ORIGIN


  Query Match           3.8%;  Score 23;  DB 71;  Length 157402;
  Best Local Similarity  100.0%;  Pred. No. 0.078;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     163 agaattatctagccccaaatgtc 185
           |||||||||||||||||||||||
Db   13008 AGAATTATCTAGCCCCAAATGTC 13030


RESULT  14
AC021862
LOCUS       AC021862   159723 bp    DNA              HTG       07-JUL-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-674E24, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
```

```
ACCESSION   AC021862
VERSION     AC021862.4  GI:8570373
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159723)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 159723)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 17, 2000 this sequence version replaced gi:7235315.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0674E24
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 76%
            Sequencing vector: plasmid; 24%
            Chemistry: Dye-primer ET; 76% of reads
            Chemistry: Dye-terminator Big Dye; 24% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 146571 bases at least Q40
            Consensus quality: 150482 bases at least Q30
            Consensus quality: 152816 bases at least Q20
            Insert size: 167000; agarose-fp
            Insert size: 158023; sum-of-contigs
            Quality coverage: 3.44 in Q20 bases; agarose-fp
            Quality coverage: 3.71 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1195: contig of 1195 bp in length
            *     1196     1295: gap of unknown length
            *     1296     3879: contig of 2584 bp in length
            *     3880     3979: gap of unknown length
            *     3980     6597: contig of 2618 bp in length
            *     6598     6697: gap of unknown length
            *     6698    10040: contig of 3343 bp in length
            *    10041    10140: gap of unknown length
            *    10141    12973: contig of 2833 bp in length
            *    12974    13073: gap of unknown length
            *    13074    15987: contig of 2914 bp in length
            *    15988    16087: gap of unknown length
            *    16088    19014: contig of 2927 bp in length
            *    19015    19114: gap of unknown length
            *    19115    22164: contig of 3050 bp in length
            *    22165    22264: gap of unknown length
            *    22265    26798: contig of 4534 bp in length
            *    26799    26898: gap of unknown length
            *    26899    34198: contig of 7300 bp in length
            *    34199    34298: gap of unknown length
            *    34299    42840: contig of 8542 bp in length
            *    42841    42940: gap of unknown length
            *    42941    48833: contig of 5893 bp in length
            *    48834    48933: gap of unknown length
            *    48934    59689: contig of 10756 bp in length
            *    59690    59789  gap of unknown length
            *    59790    72021: contig of 12232 bp in length
            *    72022    72121: gap of unknown length
            *    72122    87123: contig of 15002 bp in length
            *    87124    87223: gap of unknown length
            *    87224   106312: contig of 19089 bp in length
            *   106313   106412: gap of unknown length
            *   106413   133985: contig of 27573 bp in length
            *   133986   134085: gap of unknown length
            *   134086   159723: contig of 25638 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159723
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="RP11-674E24"
     misc_feature    1. .1195
                     /note="assembly_name:Contig7"
     misc_feature    1296. .3879
                     /note="assembly_name:Contig8"
     misc_feature    3980. .6597
                     /note="assembly_name:Contig9"
     misc_feature    6698. .10040
                     /note="assembly_name:Contig10
                     clone_end:SP6
                     vector_side:left"
     misc_feature    10141. .12973
                     /note="assembly_name:Contig11"
     misc_feature    13074. .15987
                     /note="assembly_name:Contig12"
     misc_feature    16088. .19014
                     /note="assembly_name:Contig13"
     misc_feature    19115. .22164
                     /note="assembly_name:Contig14"
     misc_feature    22265. .26798
                     /note="assembly_name:Contig15"
     misc_feature    26899. .34198
                     /note="assembly_name:Contig16"
     misc_feature    34299. .42840
                     /note="assembly_name:Contig17"
     misc_feature    42941. .48833
                     /note="assembly_name:Contig18"
     misc_feature    48934. .59689
                     /note="assembly_name:Contig19"
     misc_feature    59790. .72021
                     /note="assembly_name:Contig20"
     misc_feature    72122. .87123
                     /note="assembly_name:Contig21"
     misc_feature    87224. .106312
                     /note="assembly_name:Contig22"
     misc_feature    106413. .133985
                     /note="assembly_name:Contig23"
     misc_feature    134086. .159723
                     /note="assembly_name:Contig24"
BASE COUNT    48651 a   29918 c   29977 g   49449 t    1728 others
ORIGIN


  Query Match             3.8%;  Score 23;  DB 66;  Length 159723;
  Best Local Similarity   100.0%;  Pred. No. 0.078;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     163 agaattatctagccccaaatgtc 185
           |||||||||||||||||||||||
Db    3629 AGAATTATCTAGCCCCAAATGTC 3651


RESULT  15
AC078953
LOCUS       AC078953   171589 bp    DNA             HTG       07-JAN-2001
DEFINITION  Homo sapiens chromosome 3q clone RP11-370M22, *** SEQUENCING IN
            PROGRESS ***, 61 unordered pieces.

ACCESSION   AC078953
VERSION     AC078953.12  GI:12039127
```

KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 171589)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 171589)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (13-AUG-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jan 5, 2001 this sequence version replaced gi:11995537.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: HBSZ
Center clone name: RP11-370M22
---------- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy: 52% of reads
Chemistry: Dye-terminator Big Dye: 48% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 135345 bases at least Q40
Consensus quality: 149003 bases at least Q30
Consensus quality: 157760 bases at least Q20
Estimated insert size: 151105; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.5x in Q20 bases; sum-of-contigs estimation

---
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 9607: contig of 9607 bp in length
* 9608 9707: gap of unknown length
* 9708 15420: contig of 5713 bp in length
* 15421 15520: gap of unknown length
* 15521 21934: contig of 6414 bp in length
* 21935 22034: gap of unknown length
* 22035 28096: contig of 6062 bp in length
* 28097 28196: gap of unknown length
* 28197 33873: contig of 5677 bp in length
* 33874 33973: gap of unknown length
* 33974 39416: contig of 5443 bp in length
* 39417 39516: gap of unknown length
* 39517 44349: contig of 4833 bp in length
* 44350 44449: gap of unknown length
* 44450 48664: contig of 4215 bp in length
* 48665 48764: gap of unknown length
* 48765 53674: contig of 4910 bp in length
* 53675 53774: gap of unknown length
* 53775 58651: contig of 4877 bp in length
* 58652 58751: gap of unknown length
* 58752 63000: contig of 4249 bp in length
* 63001 63100: gap of unknown length
* 63101 66683: contig of 3583 bp in length
* 66684 66783: gap of unknown length
* 66784 69705: contig of 2922 bp in length
* 69706 69805: gap of unknown length
* 69806 73159: contig of 3354 bp in length
* 73160 73259: gap of unknown length
* 73260 76844: contig of 3585 bp in length
* 76845 76944: gap of unknown length
* 76945 80240: contig of 3296 bp in length
* 80241 80340: gap of unknown length
* 80341 83170: contig of 2830 bp in length
* 83171 83270: gap of unknown length
* 83271 85770: contig of 2500 bp in length
* 85771 85870: gap of unknown length
* 85871 88260: contig of 2390 bp in length
* 88261 88360: gap of unknown length
* 88361 91251: contig of 2891 bp in length
* 91252 91351: gap of unknown length
* 91352 95138: contig of 3787 bp in length
* 95139 95238: gap of unknown length
* 95239 97514: contig of 2276 bp in length
* 97515 97614: gap of unknown length
* 97615 99783: contig of 2169 bp in length
* 99784 99883: gap of unknown length
* 99884 102499: contig of 2616 bp in length
* 102500 102599: gap of unknown length
* 102600 105018: contig of 2419 bp in length
* 105019 105118: gap of unknown length
* 105119 107726: contig of 2608 bp in length
* 107727 107826: gap of unknown length
* 107827 111312: contig of 3486 bp in length
* 111313 111412: gap of unknown length
* 111413 114820: contig of 3408 bp in length
* 114821 114920: gap of unknown length
* 114921 117190: contig of 2270 bp in length
* 117191 117290: gap of unknown length
* 117291 119277: contig of 1987 bp in length
* 119278 119377: gap of unknown length
* 119378 120960: contig of 1583 bp in length
* 120961 121060: gap of unknown length

```
          *   121061   122181: contig of 1121 bp in length
          *   122182   122281: gap of unknown length
          *   122282   124038: contig of 1757 bp in length
          *   124039   124138: gap of unknown length
          *   124139   125666: contig of 1528 bp in length
          *   125667   125766: gap of unknown length
          *   125767   127725: contig of 1959 bp in length
          *   127726   127825: gap of unknown length
          *   127826   130376: contig of 2551 bp in length
          *   130377   130476: gap of unknown length
          *   130477   132820: contig of 2344 bp in length
          *   132821   132920: gap of unknown length
          *   132921   135499: contig of 2579 bp in length
          *   135500   135599: gap of unknown length
          *   135600   137202: contig of 1603 bp in length
          *   137203   137302: gap of unknown length
          *   137303   139112: contig of 1810 bp in length
          *   139113   139212: gap of unknown length
          *   139213   140413: contig of 1201 bp in length
          *   140414   140513: gap of unknown length
          *   140514   142609: contig of 2096 bp in length
          *   142610   142709: gap of unknown length
          *   142710   144637: contig of 1928 bp in length
          *   144638   144737: gap of unknown length
          *   144738   146091: contig of 1354 bp in length
          *   146092   146191: gap of unknown length
          *   146192   147858: contig of 1667 bp in length
          *   147859   147958: gap of unknown length
          *   147959   149631: contig of 1673 bp in length
          *   149632   149731: gap of unknown length
          *   149732   150740: contig of 1009 bp in length
          *   150741   150840: gap of unknown length
          *   150841   152348: contig of 1508 bp in length
          *   152349   152448: gap of unknown length
          *   152449   153976: contig of 1528 bp in length
          *   153977   154076: gap of unknown length
          *   154077   155379: contig of 1303 bp in length
          *   155380   155479: gap of unknown length
          *   155480   157688: contig of 2209 bp in length
          *   157689   157788: gap of unknown length
          *   157789   159085: contig of 1297 bp in length
          *   159086   159185: gap of unknown length
          *   159186   160288: contig of 1103 bp in length
          *   160289   160388: gap of unknown length
          *   160389   162315: contig of 1927 bp in length
          *   162316   162415: gap of unknown length
          *   162416   163543: contig of 1128 bp in length
          *   163544   163643: gap of unknown length
          *   163644   164773: contig of 1130 bp in length

 Query Match             3.8%;  Score 23;  DB 75;  Length 171589;
 Best Local Similarity   100.0%;  Pred. No. 0.079;
 Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    163 agaattatctagccccaaatgtc 185
          |||||||||||||||||||||||
Db   1643 AGAATTATCTAGCCCCAAATGTC 1665


Search completed: November  5, 2001, 23:44:59
Job time: 2829 sec
```

Wed Nov 7 09:21:20 2001

us-09-656-668-198.rng

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: November 5, 2001, 22:33:25 ; Search time 126.12 Seconds
(without alignments)
3017.034 Million cell updates/sec

Title: US-09-656-668-198
Perfect score: 606
Sequence: 1 tgagtttgccccttacccc.....aagcctgttcttgtctgcac 606

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	606	100.0	606	22	AAF95006	Human ovarian canc
2	82.4	13.6	80240	20	AAV83940	NC-contig derived
3	82.4	13.6	80595	20	AAV83939	HC-contig derived
4	73	12.0	2099	21	AAA27046	Human cell surface
5	73	12.0	2116	21	AAC59845	Human secreted pro
6	71	11.7	2061	20	AAZ33929	Human PRO329 nucle
7	71	11.7	2061	21	AAC78471	Human PRO329 (UNQ2
8	68.2	11.3	208	21	AAF15763	Human prostate can
9	67.4	11.1	1829	21	AAC69639	Human secreted pro
10	66.2	10.9	119950	20	AAX90201	Human yes1 gene.
11	65.8	10.9	3300	21	AAA26449	Human secreted pro
12	65.8	10.9	3300	21	AAA26449	Human secreted pro
c 13	64.8	10.7	7505	20	AAV83949	Bacterial artifici
14	64.6	10.7	1583	21	AAC79962	Human secreted pro
15	63.4	10.5	2020	21	AAC93321	Human secreted pro
16	61.2	10.1	1119	22	AAF97864	Human neuroblastom
c 17	61.2	10.1	121162	21	AAC66548	Human kinesin-like
18	60.6	10.0	11570	21	AAA95905	Human KLK-L2 gene.
c 19	60.4	10.0	6314	14	AAQ40341	Sequence of a cDNA
c 20	60	9.9	50000	21	AAA64139	Nucleotide sequenc
21	56.2	9.3	10282	19	AAV09023	Homo sapiens 20q13
22	56.2	9.3	162450	21	AAZ86967	Retinoblastoma bin
c 23	56	9.2	251	19	AAX10977	Human biallelic po
c 24	54.8	9.0	1920	21	AAC59071	Human secreted pro
c 25	54.8	9.0	1921	22	AAF32739	Human secreted pro
26	54.8	9.0	2499	21	AAA80607	Human secreted pro
27	54.8	9.0	2522	21	AAA80618	Lung cancer associ
28	54.8	9.0	2547	21	AAF18213	Human ORFX ORF2097
c 29	54.6	9.0	6439	21	AAC76542	Human KCNQ5 (KCN6q
c 30	54.2	8.9	125910	21	AAC64370	Human flavin-conta
31	53.8	8.9	25464	19	AAV57274	Human kidney amino
32	53.2	8.8	49998	20	AAX23518	Human secreted pro
c 33	53	8.7	354	21	AAC22185	Novel human polynu
34	52.8	8.7	381	22	AAF66536	Human secreted pro
35	52.8	8.7	476	21	AAC32147	Human secreted pro
c 36	52.8	8.7	693	20	AAX20444	Human secreted pro
c 37	52.6	8.7	155	21	AAC16692	Human secreted pro
c 38	52	8.6	438	20	AAX97944	Human pancreatic c
c 39	52	8.6	2423	21	AAC98795	Oligonucleotide D1
40	51.6	8.5	936	22	AAF58252	Oligonucleotide D1
41	51.6	8.5	936	22	AAF58254	Oligonucleotide D1
42	51.6	8.5	936	22	AAF58257	Oligonucleotide D2
43	51.6	8.5	936	22	AAF58259	Oligonucleotide D2
44	51.6	8.5	936	22	AAF58262	Oligonucleotide D1
45	51.6	8.5	938	22	AAF58255	

ALIGNMENTS

RESULT 1
AAF95006
ID AAF95006 standard; DNA; 606 BP.
XX
AC AAF95006;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human ovarian cancer associated coding sequence SEQ ID NO: 198.
XX
KW Human, ovarian cancer; vaccine; gene therapy; carcinoma; ds.
XX
OS Homo sapiens.
XX
PN WO200118046-A2.
XX
PD 15-MAR-2001.
XX
PF 08-SEP-2000; 2000WO-US24827.
XX
PR 10-SEP-1999; 99US-0394374.
PR 01-MAY-2000; 2000US-0561778.
PR 15-AUG-2000; 2000US-0640173.
PR 07-SEP-2000; 2000US-0656668.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Stolk JA;
XX
DR WPI; 2001-211395/21.
XX
PT Isolated polypeptides associated with ovarian carcinomas, and the
PT nucleic acids that encode them, useful for the prevention diagnosis and
PT treatment of ovarian cancers -

```
XX
PS   Claim 18; Page 189; 189pp; English.
XX
CC   The present invention provides a number of coding sequences and proteins,
CC   the over-expression of which is associated with ovarian carcinoma/cancer.
CC   These can be used in the diagnosis, treatment and prevention of ovarian
CC   cancer, optionally by gene therapy or in the form of a vaccine. The
CC   present sequence is an example of one of these sequences.
XX
SQ   Sequence 606 BP; 175 A; 147 C; 123 G; 161 T; 0 other;

  Query Match           100.0%;  Score 606;  DB 22;  Length 606;
  Best Local Similarity 100.0%;  Pred. No. 7e-187;
  Matches  606;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tgagtttgccccctttaccccatcccagtgaatatttgcaattcctaaagacgtgttttg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 tgagtttgccccctttaccccatcccagtgaatatttgcaattcctaaagacgtgttttg 60

Qy      61 attgtcacacctgggtggggaacatgctactggcatctaatgcatagagggcagtaatgc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 attgtcacacctgggtggggaacatgctactggcatctaatgcatagagggcagtaatgc 120

Qy     121 tgctaaacatctttcaacgcacaggacagagccccacaaaagagaattatctagccccaa 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 tgctaaacatctttcaacgcacaggacagagccccacaaaagagaattatctagccccaa 180

Qy     181 atgtccataacactgctgttgagaaaacctaccgcaggatcttactgggcttcataggta 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 atgtccataacactgctgttgagaaaacctaccgcaggatcttactgggcttcataggta 240

Qy     241 agcttgcctttgttctggcttctgtagatatataaaataaagacactgcccagtccctcc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 agcttgcctttgttctggcttctgtagatatataaaataaagacactgcccagtccctcc 300

Qy     301 ctcaacgtcccgagccagggctcaaggcaattccaataacagtagaatgaacactaaata 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ctcaacgtcccgagccagggctcaaggcaattccaataacagtagaatgaacactaaata 360

Qy     361 ttgatttcaaaatctcagcaactagaagaatgaccaaccatcctggttggcctgggactg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ttgatttcaaaatctcagcaactagaagaatgaccaaccatcctggttggcctgggactg 420

Qy     421 tcctagttttagcattgaaagtttcaggttccaggaaagccctcaggcctgggctgctgg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 tcctagttttagcattgaaagtttcaggttccaggaaagccctcaggcctgggctgctgg 480

Qy     481 tcaccctagcagctgagggactcttcaatacagaattagtctttgtgcactggagatgaa 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 tcaccctagcagctgagggactcttcaatacagaattagtctttgtgcactggagatgaa 540

Qy     541 tatactttaatttgtaacatgtgaaaacatctataaacatctactgaagcctgttcttgt 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 tatactttaatttgtaacatgtgaaaacatctataaacatctactgaagcctgttcttgt 600

Qy     601 ctgcac 606
          ||||||
Db     601 ctgcac 606


RESULT   2
AAV83940
ID   AAV83940 standard; DNA; 80240 BP.
XX
AC   AAV83940;
XX
DT   03-MAR-1999  (first entry)
XX
DE   NC-contig derived from mardel(10) on chromosome 10q25.2.
XX
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS   Homo sapiens.
XX
PN   WO9851790-A1.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-AU00352.
XX
PR   26-AUG-1997;   97AU-0008791.
PR   13-MAY-1997;   97AU-0006784.
XX
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI   Cancilla MR,  Choo K,  Du Sart D;
XX
DR   WPI; 1999-009773/01.
XX
PT   New isolated nucleic acid comprising neocentromere sequences from
PT   eukaryotic chromosome - used to produce replicable, segregating
PT   artificial chromosomes that can carry large amounts of DNA for gene
PT   therapy
XX
PS   Claim 9; Fig 16A; 540pp; English.
XX
CC   The present sequence represents the NC-contig derived from a mutated
CC   human chromosome 10, 10q25.2 region. The sequence contains
CC   an unusual chromosomal marker referred to as mardel(10). The
CC   mardel(10) marker is mitotically stable and contains a functional
CC   neocentromere at a location regarded as non-centromeric. This
CC   neocentromere maps to q25.2 on chromosome 10. The specification describes
CC   nucleic acid sequences derived from a eukaryotic chromosome, including a
CC   neocentromere or its functional derivative or hybrid, that are able, in
CC   a compatible cell, of replicating, acting as extra-chromosomal element
CC   and segregating during cell division. The sequences can be used to
CC   construct artificial chromosomes for use in gene therapy comprising a
CC   replicable, segregating nucleic acid that confers a specific phenotype
CC   on cells. Human artificial chromosomes can propagate in human cells and
CC   carry large amounts of DNA (e.g. therapeutic genes), and, being
CC   extra-chromosomal, they are not mutagenic. The artificial chromosomes
CC   are also useful for generation of transgenic plants and animals, in
CC   production of proteins and to make diagnostic reagents, e.g. for
CC   expression of cytokines, receptors and growth factors, or to increase
CC   the copy number of a gene in a cell. The constructs may also be
CC   used for functional and structural analysis of chromosomes.
XX
SQ   Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 8 other;

  Query Match           13.6%;  Score 82.4;  DB 20;  Length 80240;
  Best Local Similarity 63.8%;  Pred. No. 1.2e-15;
  Matches  125;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy      72 tgggtggggaacatgctactggcatctaatgcatagagggcagtaatgctgctaaacatc 131
           || ||||||     ||||||||||||||| ||  | |||  |||  |||||| ||||||||
Db   66530 tgagtgggtgggttgctactggcatctagtgggtggagaccagagatgctgttaaacatc 66589

Qy     132 tttcaacgcacaggacagagccccacaaaagagaattatctagccccaaatgtccataac 191
              ||| ||||||||||  ||| |||| | ||||||||| ||||||||| || |||
Db   66590 ccgcaaagcacaggacagtccccgacaacaaagaattatctggccccaaatatcaatagt 66649

Qy     192 actgctgttgagaaaacctaccgcaggatcttactgggcttcataggtaagcttgccttt 251
           |    ||||||||| |  |    ||  || |  |    |  |      |    |  ||
Db   66650 gccaaagttgagaaacctcattctagcttcctttcccttctacgttctaatcaactgtt 66709

Qy     252 gttctggcttctgtag 267
           |||||  |  |  |||
Db   66710 gttctttcagcattag 66725


RESULT   3
```

```
AAV83939
ID   AAV83939 standard; DNA; 80595 BP.
XX
AC   AAV83939;
XX
DT   03-MAR-1999  (first entry)
XX
DE   HC-contig derived from normal human chromosome 10q25.2 region.
XX
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS   Homo sapiens.
XX
PN   WO9851790-A1.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-AU00352.
XX
PR   26-AUG-1997;   97AU-0008791.
PR   13-MAY-1997;   97AU-0006784.
XX
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI   Cancilla MR,  Choo K,  Du Sart D;
XX
DR   WPI; 1999-009773/01.
XX
PT   New isolated nucleic acid comprising neocentromere sequences from
PT   eukaryotic chromosome - used to produce replicable, segregating
PT   artificial chromosomes that can carry large amounts of DNA for gene
PT   therapy
XX
PS   Claim 8; Fig 6; 540pp; English.
XX
CC   The present sequence represents the HC-contig derived from normal human
CC   chromosome 10, 10q25.2 region. This region can be naturally mutated to
CC   produce an unusual chromosomal marker referred to as mardel(10). The
CC   mardel(10) marker is mitotically stable and contains a functional
CC   neocentromere at a location regarded as non-centromeric. This
CC   neocentromere maps to q25.2 on chromosome 10. The specification describes
CC   nucleic acid sequences derived from a eukaryotic chromosome, including a
CC   neocentromere or its functional derivative or hybrid, that are able, in
CC   a compatible cell, of replicating, acting as extra-chromosomal element
CC   and segregating during cell division. The sequences can be used to
CC   construct artificial chromosomes for use in gene therapy comprising a
CC   replicable, segregating nucleic acid that confers a specific phenotype
CC   on cells. Human artificial chromosomes can propagate in human cells and
CC   carry large amounts of DNA (e.g. therapeutic genes), and, being
CC   extra-chromosomal, they are not mutagenic. The artificial chromosomes
CC   are also useful for generation of transgenic plants and animals, in
CC   production of proteins and to make diagnostic reagents, e.g. for
CC   expression of cytokines, receptors and growth factors, or to increase
CC   the copy number of a gene in a cell. The constructs may also be
CC   used for functional and structural analysis of chromosomes.
XX
SQ   Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 other;

  Query Match           13.6%;  Score 82.4;  DB 20;  Length 80595;
  Best Local Similarity  63.8%;  Pred. No. 1.2e-15;
  Matches  125;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy      72 tgggtggggaacatgctactggcatctaatgcatagagggcagtaatgctgctaaacatc 131
           ||  |||||     ||||||||||||||| ||  | |||  |||  |||||| ||||||||
Db   66788 tgagtgggtgggttgctactggcatctagtgggtggagaccagagatgctgttaaacatc 66847

Qy     132 tttcaacgcacaggacagagccccacaaaagagaattatctagccccaaatgtccataac 191
           |||  ||||||||||||  ||| |||| | |||||||||| |||||||||  || |||
Db   66848 ccgcaaagcacaggacagtccccgacaacaaagaattatctggccccaaatatcaatagt 66907

Qy     192 actgctgttgagaaaacctaccgcaggatcttactgggcttcataggtaagcttgccttt 251
             |   ||||||||| |  |    ||  || | |    | |      |    | ||
Db   66908 gccaaagttgagaaacctcattctagcttccttttcccttctacgttctaatcaactgtt 66967

Qy     252 gttctggcttctgtag 267
           |||||  |  |  |||
Db   66968 gttctttcagcattag 66983


RESULT   4
AAA27046
ID   AAA27046 standard; cDNA; 2099 BP.
XX
AC   AAA27046;
XX
DT   22-AUG-2000  (first entry)
XX
DE   Human cell surface receptor protein cDNA sequence #3.
XX
KW   Human; HCSRP; cytostatic; antiarthritic; antirheumatic; antiasthmatic;
KW   immunosuppressive; antiarteriosclerotic; antibacterial; antiparasitic;
KW   neuroprotective; nootropic; anticonvulsant; cancer; leukaemia;
KW   melanoma; rheumatoid arthritis; asthma; atherosclerosis; akathesia;
KW   Alzheimer's diseases; multiple sclerosis; epilepsy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             43..1119
FT                   /*tag= a
FT                   /product= "HCSRP-3"
FT
XX
PN   WO200028032-A2.
XX
PD   18-MAY-2000.
XX
PF   12-NOV-1999;   99WO-US26742.
XX
PR   12-NOV-1998;   98US-0191280.
PR   07-DEC-1998;   98US-0206647.
PR   08-MAR-1999;   99US-0123404.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tang YT,  Corley NC,  Guegler KJ,  Yue H,  Baughn MR,  Lal P;
PI   Hillman JL,  Bandman O,  Azimzai Y,  Au-Young J;
XX
DR   WPI; 2000-376546/32.
DR   P-PSDB; AAY94336.
XX
PT   New human cell surface receptor protein and polynucleotide useful for
PT   diagnosis, prevention and treatment of cancer, immune disorders,
PT   infection and neuronal disorders -
XX
PS   Claim 9; Page 89-90; 97pp; English.
XX
CC   The present sequence encodes a novel human cell surface receptor
CC   protein (HCSRP) designated HCSRP-3. The nucleotide sequence was
CC   identified in Incyte Clone 53181 from the cDNA library LNODNNOT02,
CC   which was made from RNA isolated from lymph node tissue. A number
CC   of Incyte Clones were used to assemble the consensus sequence.
CC   BLAST analysis showed that the sequence is homologous to IgG Fc
CC   receptor g583604. HCSRP and its antagonist are useful for preventing or
CC   treating disorders associated with decreased or increased expression or
CC   activity of HCSRP. Such disorders include cancers such as leukaemia and
CC   melanoma, immune disorders such as rheumatoid arthritis, asthma and
CC   atherosclerosis, bacterial and parasitic infections and neuronal
CC   disorders such as akathesia, Alzheimer's disease, multiple sclerosis and
CC   epilepsy. Polynucleotides encoding HSCRPs may be used as hybridisation
CC   probes to diagnose these conditions. Anti-HCSRP antibodies may be used
CC   as antagonists, as a targeting or delivery mechanism for bringing
CC   pharmaceutical agents into contact with cells or tissues expressing
```

CC HCSRP and for diagnosis of HCSRP-related disorders. HCSRP and its
CC catalytic or immunogenic fragments are useful for drug screening using
CC libraries of compounds.
XX
SQ Sequence 2099 BP; 528 A; 534 C; 501 G; 536 T; 0 other;

Query Match 12.0%; Score 73; DB 21; Length 2099;
Best Local Similarity 72.9%; Pred. No. 2e-13;
Matches 94; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 85 tgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacgcacag 144
Db 1523 tgctactggcatccagtaaatagaagccaggggtgccgctaaacatcctataatgcacag 1582

Qy 145 gacagagccccacaaaagagaattatctagccccaaatgtccataacactgctgttgaga 204
Db 1583 ggcagtaccccacaacgaaaaataatctggcccaaaatgtcagttgtactgagtttgaga 1642

Qy 205 aaacctacc 213
Db 1643 aaccccagc 1651

RESULT 5
AAC59845
ID AAC59845 standard; DNA; 2116 BP.
XX
AC AAC59845;
XX
DT 26-JAN-2001 (first entry)
XX
DE Human secreted protein encoding DNA clone vq24 1.
XX
KW Secreted protein; human; autoimmune disorder; multiple sclerosis; ulcer;
KW systemic lupus erythematosus; rheumatoid arthritis; anaemia; stroke;
KW haematopoiesis regulation; tissue regrowth; wound healing; haemophilia;
KW Alzheimer's disease; Parkinson's disease; Shy-drager syndrome; cancer;
KW contraceptive; infection; growth inhibition; hyperproliferative disorder;
KW psoriasis; ds.
XX
OS Homo sapiens.
XX
PN WO200055375-A1.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-US07285.
XX
PR 17-MAR-1999; 99US-0124808.
PR 17-MAR-1999; 99US-0124916.
PR 17-AUG-1999; 99US-0149639.
PR 01-OCT-1999; 99US-0157247.
PR 29-NOV-1999; 99US-0167824.
PR 15-FEB-2000; 2000US-0182711.
XX
PA (ALPH-) ALPHAGENE INC.
XX
PI Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;
XX
DR WPI; 2000-638211/61.
DR P-PSDB; AAB34744.
XX
PT Novel proteins and polypeptides useful for the treatment of e.g
PT multiple sclerosis, systemic lupus erythmatosus, rheumatoid arthritis,
PT cancer, Alzheimer's disease, Parkinson's disease, stroke, anemia and
PT ulcers
XX
PS Claim 124; Page 462; 493pp; English.
XX
CC This invention relates to 59 human secreted proteins and the nucleotide
CC sequences encoding them. Sequences AAC59788-C59846 and AAB34687-B34745
CC represent the proteins and their encoding nucleotide sequences, and
CC sequences AAB34746-B34771 represent fragments of the proteins. Probes
CC for the DNA sequences are represented by sequences AAC59847-C59596. The
CC proteins exhibit neuroprotective, dermatological, immunosuppressive,
CC antiinflammatory, antianaemic, nootropic, antiparkinsonian,
CC cerebroprotective, haemostatic, vulnerary, cytostatic, antipsoriatic,
CC antibacterial, virucide, and fungicide activity. The proteins and
CC nucleotide sequences are useful as nutritional sources or supplements
CC and in research. The proteins are useful for treating immune deficiency
CC and disorders, which may be genetic or resulting from infections,
CC autoimmune disorders such as multiple sclerosis, systemic lupus
CC erythmatosus, rheumatoid arthritis, and for treating myeloid or lymphoid
CC cell deficiencies such as anaemias by regulating haematopoiesis. The
CC proteins are also useful in compositions for bone, cartilage, tendon,
CC ligament and/or nerve tissue growth or regeneration, for wound healing,
CC tissue repair and replacement and in the treatment of wounds, incisions
CC and ulcers. Other uses include in the treatment of central and
CC peripheral nervous system and neuropathies such as Alzheimer's and
CC Parkinson's diseases and Shy-Drager syndrome, and mechanical and
CC traumatic disorders, such as spinal cord disorders, head trauma and
CC stroke. The proteins may also be used as a contraceptive, and for
CC treating coagulation disorders such as haemophilias. The protein and
CC nucleotide sequences with cadherin activity are useful for treating
CC cancer. Other uses for the protein include for inhibiting the growth,
CC infection or function of, or killing, infectious agents such as bacteria,
CC virus, fungi and other parasites, for effecting bodily characteristics
CC such as height, weight, hair colour, effecting biorhythms or cardiac
CC cycles or rhythms, effecting metabolism, catabolism, anabolism,
CC processing, utilization, storage or elimination of dietary fat, lipid,
CC protein, carbohydrate, vitamins, minerals, cofactors, effecting
CC behavioural characteristics, providing analgesic effects and for treating
CC hyperproliferative disorders such as psoriasis.
XX
SQ Sequence 2116 BP; 545 A; 531 C; 503 G; 537 T; 0 other;

Query Match 12.0%; Score 73; DB 21; Length 2116;
Best Local Similarity 72.9%; Pred. No. 2e-13;
Matches 94; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 85 tgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacgcacag 144
Db 1520 tgctactggcatccagtaaatagaagccaggggtgccgctaaacatcctataatgcacag 1579

Qy 145 gacagagccccacaaaagagaattatctagccccaaatgtccataacactgctgttgaga 204
Db 1580 ggcagtaccccacaacgaaaaataatctggcccaaaatgtcagttgtactgagtttgaga 1639

Qy 205 aaacctacc 213
Db 1640 aaccccagc 1648

RESULT 6
AAZ33929
ID AAZ33929 standard; cDNA; 2061 BP.
XX
AC AAZ33929;
XX
DT 07-DEC-1999 (first entry)
XX
DE Human PRO329 nucleotide sequence.
XX
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein; ss.
XX
OS Homo sapiens.
XX
PN WO9946281-A2.
XX
PD 16-SEP-1999.
XX
PF 08-MAR-1999; 99WO-US05028.

XX
PR 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081952.
PR 15-APR-1998; 98US-0081955.
PR 21-APR-1998; 98US-0082568.
PR 21-APR-1998; 98US-0082569.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 22-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083322.
PR 29-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083496.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 29-APR-1998; 98US-0083559.
PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.
PR 15-MAY-1998; 98US-0085573.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.
PR 15-MAY-1998; 98US-0085700.
PR 15-MAY-1998; 98US-0085704.
PR 18-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086486.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.
XX
PA (GETH) GENENTECH INC.
XX
PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
XX
DR WPI; 1999-551358/46.
DR P-PSDB; AAY41690.
XX
PT New secreted and transmembrane polypeptides and their polynucleotides,
PT useful for treating blood coagulation disorders, cancers and cellular
PT adhesion disorders -
XX
PS Claim 2; Fig 19; 530pp; English.
XX
CC The present invention describes secreted and transmembrane polypeptides
CC and their polynucleotides. The nucleotide sequences are useful as
CC sources of probes, primers, for chromosome mapping, and for generation
CC of antisense sequences. They can also be used to create transgenic
CC animals. The proteins can be used to treat a variety of diseases and
CC disorders, depending on their function. Diseases that may be treated
CC include blood coagulation disorders, cancers and cellular adhesion
CC disorders. They may also be used to raise antibodies. AAZ33891 to
CC AAZ34338, and AAY41685 to AAY41774 represent polynucleotide and
CC polypeptide sequence given in the exemplification of the present
CC invention.
XX
SQ Sequence 2061 BP; 511 A; 528 C; 491 G; 531 T; 0 other;

Query Match 11.7%; Score 71; DB 20; Length 2061;
Best Local Similarity 72.4%; Pred. No. 8.7e-13;
Matches 92; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 87 ctactggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacgcacagga 146
||||||||||| | | ||||| | ||| ||| |||||||||| | || |||||||
Db 1493 ctactggcatccagtaaatagaagccaggggtgccgctaaacatcctataatgcacaggg 1552

Qy 147 cagagccccacaaaagagaattatctagccccaaatgtccataacactgctgttgagaaaa 206
||| |||||||| | ||| |||| |||| |||||||| | |||| |||||||||
Db 1553 cagtaccccacaacgaaaaataatctggcccaaaatgtcagttgtactgagtttgagaaaa 1612

Qy 207 acctacc 213
|| | |
Db 1613 ccccagc 1619

RESULT 7
AAC78471
ID AAC78471 standard; cDNA; 2061 BP.
XX
AC AAC78471;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human PRO329 (UNQ291) nucleotide sequence SEQ ID NO:44.
XX
KW Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;

```
KW   expressed sequence tag; detection; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200053756-A2.
XX
PD   14-SEP-2000.
XX
PF   18-FEB-2000; 2000WO-US04341.
XX
PR   08-MAR-1999;   99WO-US05028.
PR   12-MAR-1999;   99US-0123957.
PR   29-MAR-1999;   99US-0126773.
PR   21-APR-1999;   99US-0130232.
PR   28-APR-1999;   99US-0131445.
PR   14-MAY-1999;   99US-0134287.
PR   23-JUN-1999;   99US-0141037.
PR   26-JUL-1999;   99US-0145698.
PR   29-OCT-1999;   99US-0162506.
PR   30-NOV-1999;   99WO-US28313.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28565.
PR   16-DEC-1999;   99WO-US30095.
PR   30-DEC-1999;   99WO-US31243.
PR   30-DEC-1999;   99WO-US31274.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
PI   Ferrara N,  Filvaroff E,  Fong S,  Gao W,  Gerber H,  Gerritsen ME;
PI   Goddard A,  Godowski PJ,  Grimaldi CJ,  Gurney AL,  Hillan KJ;
PI   Kljavin IJ,  Kuo SS,  Napier MA,  Pan J,  Paoni NF,  Roy MA;
PI   Shelton DL,  Stewart TA,  Tumas D,  Williams PM,  Wood WI;
XX
DR   WPI; 2000-611443/58.
DR   P-PSDB; AAB44246.
XX
PT   Novel PRO polypeptides and polynucleotides used in detection methods,
PT   to target bioactive molecules to specific cells, and to modulate
PT   cellular activities -
XX
PS   Claim 2; Fig 19; 636pp; English.
XX
CC   AAC78458 to AAC78599 represent polynucleotide and EST (expressed
CC   sequence tag) sequences which encode secreted or transmembrane PRO
CC   polypeptides. The PRO polynucleotides and polypeptides have cytostatic
CC   activity. The polynucleotides and polypeptides can be used for detecting
CC   the presence of PRO polypeptides in samples, for linking bioactive
CC   molecules to cells and for modulating biological activities of cells,
CC   using the polypeptides for specific targeting. The polypeptide targeting
CC   can be used to kill the target cells, e.g. for the treatment of cancers.
CC   The polypeptide pairs provide specific targeting of bioactive molecules
CC   to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
CC   the isolation of the PRO polynucleotide sequences.
XX
SQ   Sequence 2061 BP; 511 A; 528 C; 491 G; 531 T; 0 other;

  Query Match           11.7%;  Score 71;  DB 21;  Length 2061;
  Best Local Similarity  72.4%;  Pred. No. 8.7e-13;
  Matches   92;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy      87 ctactggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacgcacagga 146
           |||||||||||  | |  ||||| | |||   ||| ||||||||||  |  || |||||||
Db    1493 ctactggcatccagtaaatagaagccaggggtgccgctaaacatcctataatgcacaggg 1552

Qy     147 cagagccccacaaaagagaattatctagccccaaatgtccataacactgctgttgagaaa 206
           |||  ||||||||   | ||| |||| |||| |||||||  |   ||||  ||||||||
Db    1553 cagtaccccacaacgaaaaataatctggcccaaaatgtcagttgtactgagtttgagaaa 1612

Qy     207 acctacc 213
           || | |
Db    1613 ccccagc 1619


RESULT   8
AAF15763
ID   AAF15763 standard; cDNA; 208 BP.
XX
AC   AAF15763;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Human prostate cancer antigen nucleotide sequence SEQ ID NO:198.
XX
KW   Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW   neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW   vulnerary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW   antibacterial; gene therapy; neural; immune; reproductive; renal;
KW   gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW   wound; infectious disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055174-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05988.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587513/55.
DR   P-PSDB; AAB56560.
XX
PT   Prostate cancer associated gene sequences, referred to as prostate
PT   cancer antigens, useful for treatment, prevention, and diagnosis of
PT   disorders such as prostate cancer -
XX
PS   Claim 1; Page 736; 2338pp; English.
XX
CC   AAF15566 to AAF16505 encode the human prostate cancer associated
CC   proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC   The prostate cancer antigens can have neuroprotective, cytostatic,
CC   cardioactive, immunomodulatory, muscular, vulnerary, gastrointestinal,
CC   nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC   and can be used in gene therapy. The prostate cancer antigen
CC   polynucleotides may be used for detection of prostate cancer, chromosome
CC   identification, as chromosome markers, and for numerous other diagnostic
CC   or research purposes. The prostate cancer antigens may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC   AAB57303 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 208 BP; 63 A; 38 C; 54 G; 51 T; 2 other;

  Query Match           11.3%;  Score 68.2;  DB 21;  Length 208;
  Best Local Similarity  74.4%;  Pred. No. 2.1e-12;
  Matches   99;  Conservative    0;  Mismatches   33;  Indels    1;  Gaps    1;

Qy      79 ggaacatgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttcaac 138
           ||   ||| ||||||||||   ||  || |||  ||  ||||||||| ||||  | |
Db      31 ggggaatgttactggcatccggtgggtaaaggctagggatgctgctagacattctacggt 90

Qy     139 gcacaggacagagcccccacaaaagagaattatctagccccaaatgtccataacactgctg 198
```

```
          |||||||||| | |||||||| | ||||||||||||||| ||||||| | ||  |||  |
Db     91 gcacaggaca-acccccacaacaaagaattatctagcccaaaatgtcaacaatgctgagg 149

Qy    199 ttgagaaaaccta 211
          ||||||| ||||
Db    150 ttgagaagyccta 162
```

RESULT 9
AAC69639
ID AAC69639 standard; cDNA; 1829 BP.
XX
AC AAC69639;
XX
DT 30-JAN-2001 (first entry)
XX
DE Human secreted protein gene 43 SEQ ID NO:53.
XX
KW Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW nootropic; neuroprotective; antibacterial; virucide; fungicide;
KW opthalmalogical; autoimmune disease; hyperproliferative disorder;
KW cardiovascular disorder; cerebrovascular disorder; infection;
KW wound healing; cell proliferation; chemotaxis; food additive; ss.
XX
OS Homo sapiens.
XX
PN WO200056882-A1.
XX
PD 28-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-US06791.
XX
PR 23-MAR-1999; 99US-0125815.
PR 10-DEC-1999; 99US-0169946.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-579484/54.
DR P-PSDB; AAB38564.
XX
PT Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
XX
PS Claim 1; Page 366; 436pp; English.
XX
CC The polynucleotide sequences given in AAC69597-C69644 encode the human
CC secreted proteins given in AAB38522-B38569. Sequences given in
CC AAB38570-B38647 represent alternative proteins encoded by the genes, and
CC also represent proteins which are homologous to the gene products. The
CC secreted proteins have activities based on the tissues in which they are
CC expressed. Examples of the activities include immunosuppressive;
CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
CC vasotropic; cerebroprotective; nootropic; neuroprotective;
CC antibacterial; virucide; fungicide; and opthalmalogical. The proteins,
CC polynucleotides and/or agonists/antagonists are used to prevent, treat or
CC ameliorate medical conditions which include autoimmune diseases e.g.
CC rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the
CC breast or liver, cardiovascular disorders e.g. cardiac arrest,
CC cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous
CC system disorders e.g. Alzheimer's disease, infections caused by bacteria,
CC viruses and fungi and ocular disorders e.g. corneal infection. The
CC polypeptides can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues and in chemotaxis. The polypeptides can also be
CC used as a food additive or preservative to increase or decrease storage
CC capabilities. Sequences in AAC69597-C69644 and AAB38521 are used during
CC the isolation and characterisation of the genes of the invention.
XX
SQ Sequence 1829 BP; 545 A; 391 C; 397 G; 496 T; 0 other;

```
Query Match           11.1%;  Score 67.4;  DB 21;  Length 1829;
Best Local Similarity 67.8%;  Pred. No. 1.2e-11;
Matches  139;  Conservative    0;  Mismatches   61;  Indels    5;  Gaps    3;

Qy     21 catcccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctgggtgggg 80
          | |||||||   | | ||| |||  ||| ||||   |||||||| | |  |||||  ||
Db     89 cctcccaggagacagttgaaatgtctagagac--attttgattattatgcctggcaggac 146

Qy     81 aacatgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacgc 140
            ||  ||| |||||||||| || || ||||| |    |||||||||| | || | | |  |
Db    147 gcca--ctagtggcatctagtggatggagggtaagggtgctgctaagcgtcctcccatac 204

Qy    141 acaggacagagccccacaaaagagaattatctagccccaaatgtccataacactgctgtt 200
          |||||||||  |||| || || |||||||||| | ||||||||||  ||   |||  | |
Db    205 acaggacagcacccccacaa-agaattatccaaccccaaatgtcagtagtgctgaggct 263

Qy    201 gagaaaacctaccgcaggatcttac 225
          |||||| || ||       ||| ||
Db    264 gagaaaccccactctgctctctaac 288
```

RESULT 10
AAX90201
ID AAX90201 standard; DNA; 119950 BP.
XX
AC AAX90201;
XX
DT 23-SEP-1999 (first entry)
XX
DE Human yes1 gene.
XX
KW Human; yes1; diagnosis; neuropsychiatric disorder; BAD; schizophrenia;
KW bipolar affective disorder; attention deficit disorder;
KW schizoaffective disorder; unipolar affective disorder;
KW Huntington's disease; Parkinson's disease; manic-depression; ds.
XX
OS Homo sapiens.
XX
PN WO9935290-A1.
XX
PD 15-JUL-1999.
XX
PF 07-JAN-1999; 99WO-US00297.
XX
PR 08-JAN-1998; 98US-0003944.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Chen H, Freimer NB;
XX
DR WPI; 1999-444203/37.
DR P-PSDB; AAY24421.
XX
PT Detection of a genetic mutation in the yes1 gene, useful for
PT diagnosis of a yes1 mediated neuropsychiatric disorder in a human
XX
PS Claim 1; Fig 2; 110pp; English.
XX
CC The present invention describes a method for detecting a genetic
CC mutation in the yes1 gene for the diagnosis of a yes1 mediated
CC neuropsychiatric disorder in a human. The method comprises detecting the
CC presence or absence of a genetic mutation in the yes1 gene of the
CC subject, where the genetic mutation is a substitution, insertion or a
CC deletion and results in the production of a yes1 protein having an amino
CC acid sequence other than the wild-type yes1 amino acid sequence and the
CC presence of the genetic mutation identifies a subject that has or is at
CC risk for developing a yes1 mediated neuropsychiatric disorder. Compounds
CC that bind to the yes1 protein, alter the amount of the protein, or alter
CC the activity of the yes1 gene product, are useful for treating a yes1
CC mediated neuropsychiatric disorder. The disorders include Huntington's
CC disease, Parkinson's disease, and especially bipolar-affected disorder

```
CC   (BAD) also known as bipolar mood disorder (BP) or manic-depressive
CC   illness. The method distinguishes neuropsychiatric disorders from
CC   neurological disorders, which enables more accurate evaluation and
CC   prescription of medical treatment. The present sequence represents the
CC   human yes1 cDNA sequence.
XX
SQ   Sequence 119950 BP; 34471 A; 23730 C; 24660 G; 37033 T; 56 other;

  Query Match             10.9%;  Score 66.2;  DB 20;  Length 119950;
  Best Local Similarity   61.6%;  Pred. No. 2.7e-10;
  Matches  125;  Conservative    0;  Mismatches   73;  Indels    5;  Gaps    1;

Qy      24 cccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctgggtggggaac 83
           ||||| | |||||||  | |          ||||| |||||||||     || |||
Db   82631 cccaggggatatttggcagtatctggagacatttttggttgtcaca-----gttggggtg 82685

Qy      84 atgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacgcaca 143
           |||  ||| | |||| ||  ||| || |||  ||  || |||| |||||  |||||||||
Db   82686 ctgccgctgtcttctagtgagtagtggccagggatattggtaaatatcttaaaacgcaca 82745

Qy     144 ggacagagccccacaaaagagaattatctagccccaaatgtccataacactgctgttgag 203
             |||   |||||||| | ||||| ||| || || | |||| |||   |||  ||||
Db   82746 cagcagcctcccacaacaaagaatgatcaagtccaaggtgtcaatagtgctgaagttgga 82805

Qy     204 aaaacctaccgcaggatcttact 226
           ||| ||  |   ||  | | | |
Db   82806 aaacccagctctagactgtgaat 82828


RESULT  11
AAA26367
ID   AAA26367 standard; cDNA; 3280 BP.
XX
AC   AAA26367;
XX
DT   29-JUN-2000  (first entry)
XX
DE   Human secreted protein gene 22 SEQ ID NO:32.
XX
KW   Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW   antiHIV; antiinflammatory; nootropic; neuroprotective; antiallergic;
KW   osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
KW   antipsoriatic; cardiant; gene therapy; cancer; neurological disorder;
KW   immune disease; inflammation; blood disorder; tumour; ss.
XX
OS   Homo sapiens.
XX
PN   WO200006698-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-US17130.
XX
PR   30-JUL-1998;   98US-0094657.
PR   05-AUG-1998;   98US-0095486.
PR   06-AUG-1998;   98US-0095454.
PR   06-AUG-1998;   98US-0095455.
PR   12-AUG-1998;   98US-0096319.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Komatsoulis GA,  Rosen CA,  Ruben SM,  Duan R,  Moore PA,  Shi Y;
PI   Lafleur D,  Wei Y,  Ni J,  Florence KA,  Young PE,  Brewer LA;
PI   Soppet DR,  Endress GA,  Ebner R,  Olsen HS,  Mucenski M;
XX
DR   WPI; 2000-195282/17.
DR   P-PSDB; AAY91472.
XX
PT   New isolated human genes and the secreted polypeptides they encode,
PT   useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
XX
PS   Claim 1; Page 387-388; 634pp; English.
XX
CC   The polynucleotide sequences given in AAA26346 to AAA26458 encode the
CC   human secreted proteins given in AAY91451 to AAY91691. The human secreted
CC   proteins can have activities based on the tissues and cells they are
CC   expressed in. Examples of the activities are: cytostatic;
CC   immunosuppressive; antiHIV; antiinflammatory; nootropic; neuroprotective;
CC   antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic;
CC   antiasthma; antipsoriatic; and cardiant. The polynucleotides and their
CC   corresponding secreted proteins are useful for preventing, treating or
CC   ameliorating medical conditions, e.g. by protein or gene therapy. Also
CC   pathological conditions can be diagnosed by determining the amount of the
CC   proteins in a sample or by determining the presence of mutations in the
CC   polynucleotides. Specific uses are described for each of the
CC   polynucleotides, based on which tissues they are most highly expressed
CC   in, and include developing products for the diagnosis or treatment of
CC   cancer, tumours, neurodegenerative disorders, developmental abnormalities
CC   and foetal deficiencies, blood disorders, diseases of the immune system,
CC   autoimmune diseases, hepatic and renal disease, inflammation,
CC   allergies, Alzheimer's and behavioural disorders, schizophrenia,
CC   osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC   transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC   cardiovascular disorders, reproductive disorders, gastrointestinal
CC   disorders, respiratory disorders and metabolic disorders. The proteins
CC   or polynucleotides can also be used as food additives or preservatives.
CC   The proteins are also useful for identifying their binding partners.
CC   AAA26337 to AAA26345 and AAY91450 are sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 3280 BP; 865 A; 758 C; 762 G; 895 T; 0 other;

  Query Match             10.9%;  Score 65.8;  DB 21;  Length 3280;
  Best Local Similarity   61.3%;  Pred. No. 5.5e-11;
  Matches  106;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy      17 ccccccatcccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctgggt 76
           || |  |||||  |     |  ||||  ||   |||   ||| | |       | ||||
Db    622 cctctgtcccaaggtcatctgtcaatgactggggacacttttggttgtcataccttgggg 681

Qy      77 ggggaacatgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttca 136
           | |      ||  ||||||||||  || || |||  |||  || | ||| |||||| | ||
Db    682 gtgatgtgtgtgactggcatctggtggatggagaccagggatacagctcaacatcctaca 741

Qy     137 acgcacaggacagagccccacaaaagagaattatctagccccaaatgtccata 189
             || ||||||||   |||||||   |||| |  ||||  ||| |||||||||
Db    742 gtgcccaggacagcctcccacaatcaagaagtgcctagtgccatatgtccata 794


RESULT  12
AAA26449
ID   AAA26449 standard; cDNA; 3300 BP.
XX
AC   AAA26449;
XX
DT   29-JUN-2000  (first entry)
XX
DE   Human secreted protein gene 22 SEQ ID NO:114.
XX
KW   Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW   antiHIV; antiinflammatory; nootropic; neuroprotective; antiallergic;
KW   osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
KW   antipsoriatic; cardiant; gene therapy; cancer; neurological disorder;
KW   immune disease; inflammation; blood disorder; tumour; ss.
XX
OS   Homo sapiens.
XX
PN   WO200006698-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-US17130.
XX
```

PR 30-JUL-1998; 98US-0094657.
PR 05-AUG-1998; 98US-0095486.
PR 06-AUG-1998; 98US-0095454.
PR 06-AUG-1998; 98US-0095455.
PR 12-AUG-1998; 98US-0096319.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;
XX
DR WPI; 2000-195282/17.
DR P-PSDB; AAY91554.
XX
PT New isolated human genes and the secreted polypeptides they encode,
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders -
XX
PS Claim 1; Page 444-445; 634pp; English.
XX
CC The polynucleotide sequences given in AAA26346 to AAA26458 encode the
CC human secreted proteins given in AAY91451 to AAY91691. The human secreted
CC proteins can have activities based on the tissues and cells they are
CC expressed in. Examples of the activities are: cytostatic;
CC immunosuppressive; antiHIV; antiinflammatory; nootropic; neuroprotective;
CC antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic;
CC antiasthma; antipsoriatic; and cardiant. The polynucleotides and their
CC corresponding secreted proteins are useful for preventing, treating or
CC ameliorating medical conditions, e.g. by protein or gene therapy. Also
CC pathological conditions can be diagnosed by determining the amount of the
CC proteins in a sample or by determining the presence of mutations in the
CC polynucleotides. Specific uses are described for each of the
CC polynucleotides, based on which tissues they are most highly expressed
CC in, and include developing products for the diagnosis or treatment of
CC cancer, tumours, neurodegenerative disorders, developmental abnormalities
CC and foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, inflammation,
CC allergies, Alzheimer's and behavioural disorders, schizophrenia,
CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC cardiovascular disorders, reproductive disorders, gastrointestinal
CC disorders, respiratory disorders and metabolic disorders. The proteins
CC or polynucleotides can also be used as food additives or preservatives.
CC The proteins are also useful for identifying their binding partners.
CC AAA26337 to AAA26345 and AAY91450 are sequences used in the
CC exemplification of the present invention.
XX
SQ Sequence 3300 BP; 869 A; 768 C; 767 G; 896 T; 0 other;

Query Match 10.9%; Score 65.8; DB 21; Length 3300;
Best Local Similarity 61.3%; Pred. No. 5.5e-11;
Matches 106; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

Qy 17 cccccatcccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctgggt 76
Db 631 cctctgtcccaaggtcatctgtcaatgactggggacacttttggttgtcataccttgggg 690

Qy 77 ggggaacatgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttca 136
Db 691 gtgatgtgtgtgactggcatctggtggatggagaccagggatacagctcaacatcctaca 750

Qy 137 acgcacaggacagagccccacaaaagagaattatctagccccaaatgtccata 189
Db 751 gtgcccaggacagcctcccacaatcaagaagtgcctagtgccatatgtccata 803

RESULT 13
AAV83949/C
ID AAV83949 standard; DNA; 7505 BP.
XX
AC AAV83949;
XX
DT 03-MAR-1999 (first entry)
XX
DE Bacterial artificial chromosome (BAC)-F2 contig 9.
XX
KW Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW neocentromere; replication; extra-chromosomal element; segregation;
KW cell division; artificial chromosome; gene therapy; BAC; transgenic;
KW human artificial chromosome; bacterial artificial chromosome; ss.
XX
OS Synthetic.
XX
PN WO9851790-A1.
XX
PD 19-NOV-1998.
XX
PF 13-MAY-1998; 98WO-AU00352.
XX
PR 26-AUG-1997; 97AU-0008791.
PR 13-MAY-1997; 97AU-0006784.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Cancilla MR, Choo K, Du Sart D;
XX
DR WPI; 1999-009773/01.
XX
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy
XX
PS Claim 10; Page 222-227; 540pp; English.
XX
CC The present sequence represents a bacterial artificial chromosome (BAC)
CC contig, and exemplifies the invention. The specification describes
CC nucleic acid sequences derived from a eukaryotic chromosome, including a
CC neocentromere or its functional derivative or hybrid, that are able, in
CC a compatible cell, of replicating, acting as extra-chromosomal element
CC and segregating during cell division. The sequences can be used to
CC construct artificial chromosomes for use in gene therapy comprising a
CC replicable, segregating nucleic acid that confers a specific phenotype
CC on cells. Human artificial chromosomes can propagate in human cells and
CC carry large amounts of DNA (e.g. therapeutic genes), and, being
CC extra-chromosomal, they are not mutagenic. The artificial chromosomes
CC are also useful for generation of transgenic plants and animals, in
CC production of proteins and to make diagnostic reagents, e.g. for
CC expression of cytokines, receptors and growth factors, or to increase
CC the copy number of a gene in a cell. The constructs may also be
CC used for functional and structural analysis of chromosomes.
XX
SQ Sequence 7505 BP; 2182 A; 1718 C; 1590 G; 2015 T; 0 other;

Query Match 10.7%; Score 64.8; DB 20; Length 7505;
Best Local Similarity 72.1%; Pred. No. 1.8e-10;
Matches 98; Conservative 0; Mismatches 37; Indels 1; Gaps 1;

Qy 76 tggggaacatgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttc 135
Db 4460 TTGGGAATTTGCTACAGACATCTAATAATTAGAGGCTAGGGATATTGTTAAATGTCCTAC 4401

Qy 136 aacgcacaggacagagccccacaaaagagaattatctagccccaaatgtccataacactg 195
Db 4400 AGTGCACATGACAGACCCACATAACAGAGAATTGTATAGCCCAAAATGT-CATTAGTGTG 4342

Qy 196 ctgttgagaaaaccta 211
Db 4341 ACGTTGAGAAAATATA 4326

RESULT 14
AAC79962

ID AAC79962 standard; cDNA; 1583 BP.
XX
AC AAC79962;
XX
DT 12-FEB-2001 (first entry)
XX
DE Human secreted protein encoding cDNA for gene 15.
XX
KW Secreted protein; human; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW nootropic; neuroprotective; antibacterial; virucide; fungicide;
KW ophthalmological; vulnerary; gene therapy; treatment; autoimmune disease;
KW hyperproliferative disorder; cardiovascular disorder; ocular disorder;
KW cerebrovascular disorder; nervous system disorder; infection; skin aging;
KW wound healing; epithelial cell proliferation; transplantation; ss.
XX
OS Homo sapiens.
XX
PN WO200058357-A1.
XX
PD 05-OCT-2000.
XX
PF 23-MAR-2000; 2000WO-US07723.
XX
PR 26-MAR-1999; 99US-0126506.
PR 07-JAN-2000; 2000US-0174852.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-611704/58.
DR P-PSDB; AAB45039.
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers -
XX
PS Claim 1a; Page 349; 418pp; English.
XX
CC This invention describes novel isolated nucleic acid molecules (I)
CC encoding a human secreted proteins (II) which have immunosuppressive,
CC antiarthritic, antirheumatic, antiproliferative, cytostatic, cardiant,
CC vasotropic, cerebroprotective, nootropic, neuroprotective, antibacterial,
CC virucide, fungicide, ophthalmological and vulnerary activity and can be
CC used for gene therapy. (I) and (II) are used to prevent, treat or
CC ameliorate a medical condition in e.g. humans, mice, rabbits, goats,
CC horses, cats, dogs, chickens or sheep. (I) and (II) are also used in
CC diagnosing a pathological condition or susceptibility to a pathological
CC condition. The antibodies to (II) can also be used in alleviating
CC symptoms associated with the disorders and in diagnostic immunoassays
CC e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA).
CC Disorders which are diagnosed or treated include autoimmune diseases
CC e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms
CC of the breast or liver, cardiovascular disorders e.g. cardiac arrest,
CC cerebrovascular disorders e.g. cerebral ischemia, angiogenesis, nervous
CC system disorders e.g. Alzheimer's disease, infections caused by bacteria,
CC viruses and fungi and ocular disorders e.g. corneal infection. The
CC polypeptides can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues and in chemotaxis. The polypeptides can also be
CC used as a food additive or preservative to increase or decrease storage
CC capabilities.
XX
SQ Sequence 1583 BP; 427 A; 370 C; 316 G; 458 T; 12 other;

Query Match 10.7%; Score 64.6; DB 21; Length 1583;
Best Local Similarity 64.5%; Pred. No. 9.1e-11;
Matches 91; Conservative 3; Mismatches 47; Indels 0; Gaps 0;

Qy 73 gggtggggaacatgctactggcatctaatgcatagagggcagtaatgctgctaaacatct 132

Db 918 gaggggggtgaggtactactggcatcttgtgggtagagagtasggatgctgtttaacatcc 977

Qy 133 ttcaacgcacaggacagagccccacaaaagagaattatctagccccaaatgtccataaca 192

Db 978 tgcaatgcacaggacaggccctgacmacmaacagttatctggccccagatgccaaaagtg 1037

Qy 193 ctgctgttgagaaaacctacc 213

Db 1038 ccagagttaaagaactttgcc 1058

RESULT 15
AAC93321
ID AAC93321 standard; cDNA; 2020 BP.
XX
AC AAC93321;
XX
DT 16-FEB-2001 (first entry)
XX
DE Human secreted protein cDNA sequence #12.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
XX
OS Homo sapiens.
XX
PN WO200058495-A1.
XX
PD 05-OCT-2000.
XX
PF 23-MAR-2000; 2000WO-US07661.
XX
PR 26-MAR-1999; 99US-0126504.
PR 07-JAN-2000; 2000US-0174847.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-611720/58.
DR P-PSDB; AAB51391.
XX
PT New nucleic acid molecules encoding 45 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 1; Pages 331-332; 410pp; English.
XX
CC The invention relate to the isolation of genes AAC93310-C93354 encoding
CC 45 human secreted proteins AAB51380-B51423. The genes can be used to
CC generate fusion proteins by linking to the gene for the human
CC immunoglobulin G Fc portion (AAC93301) for increasing the stability of
CC the fusion protein as compared to the human protein only. The genes and
CC proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
XX
SQ Sequence 2020 BP; 538 A; 422 C; 403 G; 641 T; 16 other;

Query Match 10.5%; Score 63.4; DB 21; Length 2020;
Best Local Similarity 66.3%; Pred. No. 2.5e-10;
Matches 108; Conservative 0; Mismatches 51; Indels 4; Gaps 1;

```
Qy      57 tttgattgtcacacctgggtggggaacatgctactggcatctaatgcatagagggcagta 116
           |||| |||   || ||||||||||    ||||| | ||||||  ||  || ||| ||| |
Db    1814 tttggttggtccatctgggtgggg----tgctattagcatctcttgagtaaaggccagga 1869

Qy     117 atgctgctaaacatctttcaacgcacaggacagagccccacaaaagagaattatctagcc 176
           || ||||||||||  | |||  ||||||| ||  ||| ||||   | ||||||| || |
Db    1870 atactgctaaacattctgcaatacacaggatagccccctacaatgaataattatccagac 1929

Qy     177 ccaaatgtccataacactgctgttgagaaaacctaccgcagga 219
           ||| ||| | | |       |||| ||||   | |   || |
Db    1930 ccacatggcaagagtgacaaggttgtgaaactttgctctagaa 1972
```

Search completed: November 5, 2001, 23:00:36
Job time: 1631 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: November 5, 2001, 23:00:40 ; Search time 126.05 Seconds
(without alignments)
3018.710 Million cell updates/sec

Title: US-09-656-668-198
Perfect score: 606
Sequence: 1 tgagtttgccccccttacccc.....aagcctgttcttgtctgcac 606

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		606	100.0	606	22	AAF95006	Human ovarian canc
c 2		21	3.5	51	19	AAV36272	Probe used to isol
c 3		19	3.1	2769	21	AAC76385	Human ORFX ORF1940
4		18	3.0	1427	11	AAQ06902	Sequence encoding
5		18	3.0	1658	21	AAF21367	Human low adenosin
6		18	3.0	1658	21	AAA35245	Human adenosine re
7		18	3.0	1674	11	AAQ06901	Sequence encoding
c 8		18	3.0	2582	18	AAT87428	Clone J422. Homo
c 9		18	3.0	2635	19	AAV15587	Human B-cell activ
c 10		18	3.0	2697	20	AAV83816	cDNA encoding a B
c 11		18	3.0	2775	18	AAT72271	Human B cell surfa
c 12		18	3.0	2775	21	AAA08399	Human RP105 nucleo
13		18	3.0	3191	21	AAF21368	Human low adenosin
14		18	3.0	3191	21	AAA35246	Human adenosine re
15		18	3.0	1830121	17	AAT42063	Haemophilus influe
c 16		17	2.8	268	21	AAC33704	Arabidopsis thalia
17		17	2.8	352	20	AAV86186	EST clone J638. H
c 18		17	2.8	394	16	AAT26421	Human gene signatu
c 19		17	2.8	396	20	AAV87210	EST clone BN99. H
20		17	2.8	497	21	AAC53869	Arabidopsis thalia
21		17	2.8	542	21	AAC53133	Arabidopsis thalia
c 22		17	2.8	670	21	AAZ65305	Human secreted pro
c 23		17	2.8	776	18	AAT61702	Rat probasin encod
c 24		17	2.8	1329	18	AAT67996	H. pylori cytoplas
25		17	2.8	1383	18	AAT66164	Mouse interleukin-
26		17	2.8	1551	21	AAC81117	Human secreted pro
c 27		17	2.8	1680	22	AAF23411	Human SEC2 DNA seq
28		17	2.8	2364	21	AAZ55926	cDNA encoding Xeno
c 29		17	2.8	2601	20	AAX90036	Fervidobacterium p
30		17	2.8	5921	20	AAZ32700	Human glycophospha
31		17	2.8	8625	21	AAC75834	Human ORFX ORF1389
c 32		17	2.8	15512	19	AAV59470	Human ryanodin rec
c 33		17	2.8	43795	21	AAZ92583	Human DAZ genomic
34		17	2.8	240825	22	AAF24497	Human PG-3 gene.
c 35		16	2.6	166	21	AAC25754	Human secreted pro
36		16	2.6	264	21	AAC30995	Human secreted pro
37		16	2.6	269	21	AAZ93797	Sequence encoding
c 38		16	2.6	317	21	AAC29682	Human secreted pro
c 39		16	2.6	371	14	AAQ39834	Expressed Sequence
c 40		16	2.6	371	14	AAQ59246	Human brain Expres
41		16	2.6	375	21	AAC26100	Human secreted pro
42		16	2.6	378	21	AAA78073	cDNA encoding huma
43		16	2.6	401	21	AAC03103	Human secreted pro
44		16	2.6	422	21	AAZ80244	Human colon cancer
45		16	2.6	484	21	AAC01423	Human secreted pro

ALIGNMENTS

RESULT 1
AAF95006
ID AAF95006 standard; DNA; 606 BP.
XX
AC AAF95006;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human ovarian cancer associated coding sequence SEQ ID NO: 198.
XX
KW Human, ovarian cancer; vaccine; gene therapy; carcinoma; ds.
XX
OS Homo sapiens.
XX
PN WO200118046-A2.
XX
PD 15-MAR-2001.
XX
PF 08-SEP-2000; 2000WO-US24827.
XX
PR 10-SEP-1999; 99US-0394374.
PR 01-MAY-2000; 2000US-0561778.
PR 15-AUG-2000; 2000US-0640173.
PR 07-SEP-2000; 2000US-0656668.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Stolk JA;
XX
DR WPI; 2001-211395/21.
XX
PT Isolated polypeptides associated with ovarian carcinomas, and the
PT nucleic acids that encode them, useful for the prevention diagnosis and
PT treatment of ovarian cancers -

XX
PS Claim 18; Page 189; 189pp; English.
XX
CC The present invention provides a number of coding sequences and proteins,
CC the over-expression of which is associated with ovarian carcinoma/cancer.
CC These can be used in the diagnosis, treatment and prevention of ovarian
CC cancer, optionally by gene therapy or in the form of a vaccine. The
CC present sequence is an example of one of these sequences.
XX
SQ Sequence 606 BP; 175 A; 147 C; 123 G; 161 T; 0 other;

Query Match 100.0%; Score 606; DB 22; Length 606;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 606; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 tgagtttgccccttacccccatcccagtgaatatttgcaattcctaaagacgtgttttg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 tgagtttgccccttacccccatcccagtgaatatttgcaattcctaaagacgtgttttg 60

Qy     61 attgtcacacctgggtggggaacatgctactggcatctaatgcatagagggcagtaatgc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 attgtcacacctgggtggggaacatgctactggcatctaatgcatagagggcagtaatgc 120

Qy    121 tgctaaacatctttcaacgcacaggacagagccccacaaaagagaattatctagccccaa 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 tgctaaacatctttcaacgcacaggacagagccccacaaaagagaattatctagccccaa 180

Qy    181 atgtccataacactgctgttgagaaaacctaccgcaggatcttactgggcttcataggta 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 atgtccataacactgctgttgagaaaacctaccgcaggatcttactgggcttcataggta 240

Qy    241 agcttgcctttgttctggcttctgtagatatataaaataaagacactgcccagtccctcc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 agcttgcctttgttctggcttctgtagatatataaaataaagacactgcccagtccctcc 300

Qy    301 ctcaacgtcccgagccagggctcaaggcaattccaataacagtagaatgaacactaaata 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ctcaacgtcccgagccagggctcaaggcaattccaataacagtagaatgaacactaaata 360

Qy    361 ttgatttcaaaatctcagcaactagaagaatgaccaaccatcctggttggcctgggactg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ttgatttcaaaatctcagcaactagaagaatgaccaaccatcctggttggcctgggactg 420

Qy    421 tcctagttttagcattgaaagtttcaggttccaggaaagccctcaggcctgggctgctgg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 tcctagttttagcattgaaagtttcaggttccaggaaagccctcaggcctgggctgctgg 480

Qy    481 tcaccctagcagctgagggactcttcaatacagaattagtctttgtgcactggagatgaa 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 tcaccctagcagctgagggactcttcaatacagaattagtctttgtgcactggagatgaa 540

Qy    541 tatactttaatttgtaacatgtgaaaacatctataaacatctactgaagcctgttcttgt 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 tatactttaatttgtaacatgtgaaaacatctataaacatctactgaagcctgttcttgt 600

Qy    601 ctgcac 606
          ||||||
Db    601 ctgcac 606
```

RESULT 2
AAV36272/c
ID AAV36272 standard; DNA; 51 BP.
XX
AC AAV36272;
XX
DT 08-SEP-1998 (first entry)
XX
DE Probe used to isolate DNA encoding Elongin A homologue of C. elegans.
XX
KW Elongin A; rat; elongation activity; stimulation; RNA polymerase II;
KW binding; Elongin B; Elongin C; regulation; transcription activity;
KW RNA polymerase; in vitro transcriptional assay; probe; ss.
XX
OS Synthetic.
OS Caenorhabditis elegans.
XX
PN WO9814574-A1.
XX
PD 09-APR-1998.
XX
PF 02-OCT-1997; 97WO-US17992.
XX
PR 04-OCT-1996; 96US-0725459.
XX
PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.
XX
PI Conaway JW, Conaway RC;
XX
DR WPI; 1998-286420/25.
XX
PT Isolated Elongin A and C and fragments - used for regulating
PT transcriptional activation of RNA polymerase, particularly for use
PT in in vitro assays and systems
XX
PS Example 7; Page 44; 185pp; English.
XX
CC The present sequence represents a probe used to isolate DNA encoding an
CC Elongin A homologue of Caenorhabditis elegans. Elongin A stimulates
CC the elongation activity of RNA polymerase II. Fragments of the rat
CC Elongin A (AAW60772-76) have been found to each possess 50% or more of
CC the elongation activation activity exhibited by the wild type protein.
CC The Elongin A sequence critical for binding to Elongins B and C is also
CC found within these fragments. The Elongin polypeptides and fragments can
CC be used for regulating the transcriptional activity of RNA polymerase.
CC They can be used in in vitro transcriptional assays or systems.
XX
SQ Sequence 51 BP; 13 A; 9 C; 17 G; 12 T; 0 other;

Query Match 3.5%; Score 21; DB 19; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    579 atctactgaagcctgttcttg 599
          |||||||||||||||||||||
Db     44 ATCTACTGAAGCCTGTTCTTG 24
```

RESULT 3
AAC76385/c
ID AAC76385 standard; cDNA; 2769 BP.
XX
AC AAC76385;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1940 polynucleotide sequence SEQ ID NO:3879.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.

XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR P-PSDB; AAB42176.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 3025-3027; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 2769 BP; 604 A; 789 C; 772 G; 603 T; 1 other;

Query Match 3.1%; Score 19; DB 21; Length 2769;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 425 agttttagcattgaaagtt 443
||||||||||||||||||
Db 1419 AGTTTTAGCATTGAAAGTT 1401

RESULT 4
AAQ06902
ID AAQ06902 standard; DNA; 1427 BP.
XX
AC AAQ06902;
XX
DT 04-MAR-1991 (first entry)
XX
DE Sequence encoding human IL-7 receptor clone H6.
XX
KW Interleukin-7 receptor; immune response; pre-B cell growth factor;
KW Lymphopoietin; ds.
XX
OS Homo sapiens.

XX
FH Key Location/Qualifiers
FT CDS 33..818
FT /*tag= a
FT sig_peptide 33..92
FT /*tag= b
FT mat_peptide 93..818
FT /*tag= c
XX
PN EP403114-A.
XX
PD 19-DEC-1990.
XX
PF 31-MAY-1990; 90EP-0305928.
XX
PR 14-MAR-1990; 90US-0493588.
PR 15-JUN-1989; 89US-0366910.
PR 13-OCT-1989; 89US-0421201.
XX
PA (IMMU-) IMMUNEX CORP.
XX
PI Park LS, Goodwin RG;
XX
DR WPI; 1990-377843/51.
DR P-PSDB; AAR08330.
XX
PT Mammalian interleukin-7 receptor DNA, protein and analogues -
PT used in therapy, diagnosis, assay and antibody production
XX
PS Claim 4; Fig 3; 28pp; English.
XX
CC IL-7R gene product may be used in immunoassay or for affinity
CC purification eg. IL-7R, IL-1 or IL-1R. May also be useful in
CC regulation of immune response specifically of IL-7 in mammals
CC esp. humans.
XX
SQ Sequence 1427 BP; 438 A; 313 C; 313 G; 362 T; 1 other;

Query Match 3.0%; Score 18; DB 11; Length 1427;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 374 ctcagcaactagaagaat 391
||||||||||||||||||
Db 924 ctcagcaactagaagaat 941

RESULT 5
AAF21367
ID AAF21367 standard; DNA; 1658 BP.
XX
AC AAF21367;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2934.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX

PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1370; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 1658 BP; 493 A; 364 C; 363 G; 438 T; 0 other;

Query Match 3.0%; Score 18; DB 21; Length 1658;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 374 ctcagcaactagaagaat 391
||||||||||||||||||
Db 1008 ctcagcaactagaagaat 1025

RESULT 6
AAA35245
ID AAA35245 standard; DNA; 1658 BP.
XX
AC AAA35245;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:119.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 1284-1285; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 1658 BP; 493 A; 364 C; 363 G; 438 T; 0 other;

Query Match 3.0%; Score 18; DB 21; Length 1658;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 374 ctcagcaactagaagaat 391
||||||||||||||||||
Db 1008 ctcagcaactagaagaat 1025

RESULT 7
AAQ06901
ID AAQ06901 standard; cDNA; 1674 BP.
XX
AC AAQ06901;

```
XX
DT   04-MAR-1991  (first entry)
XX
DE   Sequence encoding human IL-7 receptor clone 1120.
XX
KW   Interleukin-7 receptor; immune response; pre-B cell growth factor;
KW   Lymphopoietin; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             23..1399
FT                   /*tag=  a
FT   sig_peptide     23..82
FT                   /*tag=  b
FT   mat_peptide     83..1399
FT                   /*tag=  c
FT   misc_RNA        740..814
FT                   /*tag=  d
FT                   /label= Encodes transmembrane region
FT
XX
PN   EP403114-A.
XX
PD   19-DEC-1990.
XX
PF   31-MAY-1990;   90EP-0305928.
XX
PR   14-MAR-1990;   90US-0493588.
PR   15-JUN-1989;   89US-0366910.
PR   13-OCT-1989;   89US-0421201.
XX
PA   (IMMU-) IMMUNEX CORP.
XX
PI   Park LS,  Goodwin RG;
XX
DR   WPI; 1990-377843/51.
DR   P-PSDB; AAR08329.
XX
PT   Mammalian interleukin-7 receptor DNA, protein and analogues -
PT   used in therapy, diagnosis, assay and antibody production
XX
PS   Claim 4; Fig 2; 28pp; English.
XX
CC   IL-7R gene product may be used in immunoassay or for affinity
CC   purification eg. IL-7R, IL-1 or IL-1R. May also be useful in
CC   regulation of immune response specifically of IL-7 in mammals
CC   esp. humans.
XX
SQ   Sequence 1674 BP; 509 A; 366 C; 363 G; 436 T; 0 other;


  Query Match             3.0%;  Score 18;  DB 11;  Length 1674;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     374 ctcagcaactagaagaat 391
           ||||||||||||||||||
Db    1008 ctcagcaactagaagaat 1025


RESULT   8
AAT87428/c
ID   AAT87428 standard; cDNA; 2582 BP.
XX
AC   AAT87428;
XX
DT   29-DEC-1997  (first entry)
XX
DE   Clone J422.
XX
KW   J5; J422; L105; H174-10; H174-43; B18; cytokine; PBMC;
KW   peripheral blood mononuclear cell; disintegrin; metallo-protein;
KW   Drosophila; leucine-rich repeat; monocyte; chemoattractant;
KW   IP-10; CRG-2; CTLA-8; herpesvirus; Saimiri; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             52..2037
FT                   /*tag=  a
FT
XX
PN   WO9707198-A2.
XX
PD   27-FEB-1997.
XX
PF   08-AUG-1996;   96WO-US12897.
XX
PR   08-AUG-1996;   96WO-US12897.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Carlin M,  Jacobs K,  Kelleher K,  Mccoy JM;
XX
DR   WPI; 1997-165283/15.
DR   P-PSDB; AAW28510.
XX
PT   Polynucleotide(s) encoding proteins for treating, preventing and
PT   ameliorating medical conditions - obtained from human activated
PT   peripheral blood mononuclear cell, and murine adult thymus libraries
XX
PS   Claim 11; Page 39-42; 61pp; English.
XX
CC   This sequence was isolated from a human activated PBMC library using
CC   a trap selecting for nucleotides encoding secreted proteins, and
CC   encodes a protein having homology with Drosophila leucine-rich
CC   repeat proteins.
XX
SQ   Sequence 2582 BP; 726 A; 653 C; 508 G; 695 T; 0 other;


  Query Match             3.0%;  Score 18;  DB 18;  Length 2582;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     286 ctgcccagtccctccctc 303
           ||||||||||||||||||
Db    2328 CTGCCCAGTCCCTCCCTC 2311


RESULT   9
AAV15587/c
ID   AAV15587 standard; cDNA; 2635 BP.
XX
AC   AAV15587;
XX
DT   22-MAY-1998  (first entry)
XX
DE   Human B-cell activation and survival antigen-1 cDNA.
XX
KW   Human; B-cell activation and survival antigen-1; BAS-1; diagnosis;
KW   treatment; abnormal B-cell response; abnormal proliferation;
KW   cancerous condition; degenerative condition; autoimmune disease;
KW   allergic response; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             97..2082
FT                   /*tag=  a
FT                   /product= BAS-1
FT
XX
PN   WO9744452-A1.
XX
PD   27-NOV-1997.
XX
PF   15-MAY-1997;   97WO-US07648.
```

XX
PR 17-MAY-1996; 96US-0649553.
XX
PA (SCHE) SCHERING CORP.
XX
PI Banchereau J, Fugier-Vivier I, Liu Y;
XX
DR WPI; 1998-018507/02.
DR P-PSDB; AAW47274.
XX
PT Novel human B-cell Activation and Survival antigen-1 - useful in the
PT diagnosis and treatment of abnormal B-cell responses e.g. cancerous
PT conditions or auto:immune diseases
XX
PS Claim 2; Pages 57-61; 74pp; English.
XX
CC The present sequence encodes human B-cell activation and survival
CC antigen-1 (BAS-1), which can be used in the diagnosis and treatment
CC of abnormal B-cell responses, e.g. abnormal proliferation such as
CC cancerous conditions, degenerative conditions, autoimmune diseases
CC or allergic responses.
XX
SQ Sequence 2635 BP; 736 A; 667 C; 522 G; 710 T; 0 other;

Query Match 3.0%; Score 18; DB 19; Length 2635;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 286 ctgcccagtccctccctc 303
||||||||||||||||||
Db 2373 CTGCCCAGTCCCTCCCTC 2356

RESULT 10
AAV83816/c
ID AAV83816 standard; cDNA; 2697 BP.
XX
AC AAV83816;
XX
DT 03-MAR-1999 (first entry)
XX
DE cDNA encoding a B cell surface protein.
XX
KW B cell surface protein; growth; inhibition; induction; B cell apoptosis;
KW autoimmune disease; allergic disease; asthma; atopic dermatitis; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 143..2128
FT /*tag= a
FT sig_peptide 143..203
FT /*tag= b
FT mat_peptide 204..2125
FT /*tag= c
XX
PN JP10313870-A.
XX
PD 02-DEC-1998.
XX
PF 22-MAY-1997; 97JP-0132592.
XX
PR 22-MAY-1997; 97JP-0132592.
XX
PA (MITU) MITSUBISHI CHEM CORP.
XX
DR WPI; 1999-084647/08.
DR P-PSDB; AAW87556.
XX
PT New DNA sequence encoding B cell surface protein - useful as
PT diagnostic agent and for treating auto-immune diseases, allergic
PT diseases, asthma and atopic dermatitis
XX
PS Claim 2; Page 10-13; 15pp; Japanese.
XX
CC The present sequence encodes a B cell surface protein. The protein
CC promotes growth of B cells as well as inhibition and induction of B
CC cell apoptosis. The protein and the DNA are useful in the treatment
CC and as a diagnostic agent for B-cell antibody production related
CC diseases e.g. autoimmune diseases, allergic diseases, asthma and
CC atopic dermatitis.
XX
SQ Sequence 2697 BP; 757 A; 680 C; 533 G; 727 T; 0 other;

Query Match 3.0%; Score 18; DB 20; Length 2697;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 286 ctgcccagtccctccctc 303
||||||||||||||||||
Db 2419 CTGCCCAGTCCCTCCCTC 2402

RESULT 11
AAT72271/c
ID AAT72271 standard; cDNA; 2775 BP.
XX
AC AAT72271;
XX
DT 30-SEP-1997 (first entry)
XX
DE Human B cell surface antigen cDNA.
XX
KW B cell surface antigen; gene vaccine; apoptosis; dexamethasone;
KW irradiation; ss.
XX
OS Homo sapiens.
XX
PN WO9714813-A1.
XX
PD 24-APR-1997.
XX
PF 16-OCT-1996; 96WO-US16527.
XX
PR 19-OCT-1995; 95US-0005642.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Marshall L, Roshak A;
XX
DR WPI; 1997-245128/22.
XX
PT cDNA clone encoding human B cell surface antigen - used in gene
PT vaccine and to prevent apoptosis resulting from dexamethasone
PT treatment and irradiation
XX
PS Claim 1; Page 16-17; 22pp; English.
XX
CC A cDNA clone (AAT72271) codes for a novel human B cell surface
CC antigen that is involved in B cell proliferation and protection from
CC dexamethasone- and irradiation-induced apoptosis. A plasmid
CC containing the cDNA clone was deposited as ATCC 69902. To isolate
CC the clone, cDNA libraries were constructed in the Stratagene
CC UniZapXR lambda phage vector and immediately amplified. Plasmid
CC pBluescriptSK+ (contained within UniZap phage) was excised by co-
CC infection of XL-1 Blue MRF' cells with helper phage and an aliquot
CC of the amplified library. The resulting phagemid stock was used to
CC infect SOLR cells which were then plated on LB+ ampicillin plates.
CC DNA from individual colonies was sequenced. The insert is in the
CC EcoRI(5')/XhoI(3') sites of Bluescript SK+. The isolated clone can
CC be used to produce recombinant antigen, in methods for detecting
CC mutated forms of the human gene antigen, and in the development of
CC genetic vaccines.
XX

SQ Sequence 2775 BP; 792 A; 694 C; 565 G; 724 T; 0 other;

Query Match 3.0%; Score 18; DB 18; Length 2775;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 286 ctgcccagtccctccctc 303
||||||||||||||||||
Db 2417 CTGCCCAGTCCCTCCCTC 2400

RESULT 12
AAA08399/c
ID AAA08399 standard; DNA; 2775 BP.
XX
AC AAA08399;
XX
DT 12-JUL-2000 (first entry)
XX
DE Human RP105 nucleotide sequence SEQ ID NO:2.
XX
KW Human; RP105; agonist; antagonist; allergy; allergic disease;
KW identification; antiasthmatic; dermatological; antiinflammatory;
KW cytostatic; immunosuppressive; neuroprotective; antianaemic;
KW antidiabetic; asthma; atopic dermatitis; B cell neoplasm;
KW chronic lymphocyte leukaemia; hairy cell leukaemia; myeloma;
KW prolymphocytic leukaemia; autoimmune disease; rheumatoid arthritis;
KW systemic lupus erythematosus; multiple sclerosis; diabetes;
KW acquired haemolytic anaemia; ds.
XX
OS Homo sapiens.
XX
PN WO200012130-A1.
XX
PD 09-MAR-2000.
XX
PF 24-AUG-1999; 99WO-US19336.
XX
PR 27-AUG-1998; 98US-0098030.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
XX
PI Harrop JA, Holmes SD, Roshak AK;
XX
DR WPI; 2000-375617/32.
DR P-PSDB; AAY82527.
XX
PT New anti-murine B cell antigen RP105 antibody for treating or
PT preventing allergic disease such as asthma, B cell neoplasms, for
PT example leukemias, and auto immune diseases such as systemic lupus
PT erythematosus -
XX
PS Example; Page 26-28; 30pp; English.
XX
CC The present invention describes an anti-murine B cell antigen RP105
CC antibody (I). The present sequence encodes human RP105, which is used
CC in the exemplification of the present invention. (I) has antiasthmatic,
CC dermatological, antiinflammatory, cytostatic, immunosuppressive,
CC neuroprotective, antianaemic and antidiabetic activities. (I) can be
CC used for in the treatment and prevention of allergic disease, including
CC asthma and atopic dermatitis, B cell neoplasms, including chronic
CC lymphocyte leukaemia, hairy cell leukaemia, prolymphocytic leukaemia,
CC myelomas, autoimmune diseases such as systemic lupus erythematosus,
CC rheumatoid arthritis, multiple sclerosis, acquired haemolytic anaemia
CC and diabetes.
XX
SQ Sequence 2775 BP; 792 A; 694 C; 565 G; 724 T; 0 other;

Query Match 3.0%; Score 18; DB 21; Length 2775;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 286 ctgcccagtccctccctc 303
||||||||||||||||||
Db 2417 CTGCCCAGTCCCTCCCTC 2400

RESULT 13
AAF21368
ID AAF21368 standard; DNA; 3191 BP.
XX
AC AAF21368;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2935.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1370-1371; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,

CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 3191 BP; 947 A; 674 C; 663 G; 907 T; 0 other;

Query Match 3.0%; Score 18; DB 21; Length 3191;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 374 ctcagcaactagaagaat 391
||||||||||||||||||
Db 2541 ctcagcaactagaagaat 2558

RESULT 14
AAA35246
ID AAA35246 standard; DNA; 3191 BP.
XX
AC AAA35246;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:120.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers
XX
PS Disclosure; Page 1285-1286; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 3191 BP; 947 A; 674 C; 663 G; 907 T; 0 other;

Query Match 3.0%; Score 18; DB 21; Length 3191;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 374 ctcagcaactagaagaat 391
||||||||||||||||||
Db 2541 ctcagcaactagaagaat 2558

RESULT 15
AAT42063
ID AAT42063 standard; DNA; 1830121 BP.
XX
AC AAT42063;
XX
DT 14-SEP-1999 (first entry)
XX
DE Haemophilus influenzae complete genome sequence.
XX
KW Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW expression modulating fragment; regulation; gene expression; vector;
KW organism; open reading frame; ORF; ds.
XX
OS Haemophilus influenzae.
XX
PN WO9633276-A1.
XX
PD 24-OCT-1996.
XX
PF 22-APR-1996; 96WO-US05320.
XX
PR 07-JUN-1995; 95US-0487429.
PR 21-APR-1995; 95US-0426787.
PR 07-JUN-1995; 95US-0476102.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UYJO) UNIV JOHNS HOPKINS.
XX
PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
XX
DR WPI; 1996-485782/48.
XX
PT Haemophilus influenzae Rd genome recorded on computer readable
PT medium - useful for identifying commercially important nucleic acid
PT fragments by homology searching
XX
PS Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
CC This sequence represents the complete genome sequence of the bacterium
CC Haemophilus influenzae strain Rd. The invention relates to a computer
CC readable medium (CRM) having recorded upon it the complete H.influenzae
CC nucleotide sequence (I), a representative fragment of (I) or a nucleotide
CC sequence at least 99% identical to (I). By providing the full-length
CC genomic sequence in a computer readable form, it is possible to identify
CC commercially important nucleic acid fragments and expression modulating

CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC regulate the expression of a nucleic acid molecule. Vectors and altered
CC organisms comprising the predicted ORFS can be used to produce any of the
CC polypeptide fragments of the H. influenzae Rd genome.
XX
SQ Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

Query Match 3.0%; Score 18; DB 17; Length 1830121;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 580 tctactgaagcctgttct 597
||||||||||||||||||
Db 408370 tctactgaagcctgttct 408387

Search completed: November 5,. 2001, 23:48:56
Job time: 2896 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: November 5, 2001, 22:16:40 ; Search time 87.77 Seconds
(without alignments)
1563.697 Million cell updates/sec

Title: US-09-656-668-198
Perfect score: 606
Sequence: 1 tgagtttgccccttacccc.....aagcctgttcttgtctgcac 606

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	82.4	13.6	80246	4	US-09-078-294-4	Sequence 4, Appli
	2	82.4	13.6	80595	4	US-09-078-294-3	Sequence 3, Appli
c	3	64.8	10.7	7505	4	US-09-078-294-13	Sequence 13, Appl
c	4	60.4	10.0	6314	1	US-08-211-430-1	Sequence 1, Appli
	5	46.8	7.7	279	3	US-09-157-177-120	Sequence 120, App
	6	41	6.8	1750	4	US-09-276-531-34	Sequence 34, Appl
c	7	38.6	6.4	11288	4	US-08-646-301A-1	Sequence 1, Appli
c	8	38.6	6.4	11288	4	US-08-481-968A-4	Sequence 4, Appli
c	9	36.6	6.0	3176	2	US-08-910-733-17	Sequence 17, Appl
c	10	36.6	6.0	3176	2	US-08-910-884-17	Sequence 17, Appl
	11	35.6	5.9	11531	1	US-08-068-945A-1	Sequence 1, Appli
	12	35.6	5.9	11531	1	US-08-442-806-1	Sequence 1, Appli
	13	35.4	5.8	1481	3	US-09-429-323-3	Sequence 3, Appli
	14	35.4	5.8	3073	1	US-07-688-352C-31	Sequence 31, Appl
	15	35.4	5.8	3073	2	US-08-474-379C-31	Sequence 31, Appl
	16	35.4	5.8	3073	3	US-09-146-249A-31	Sequence 31, Appl
	17	35.4	5.8	3073	3	US-08-206-188B-31	Sequence 31, Appl
	18	35.4	5.8	3073	5	PCT-US91-02714-30	Sequence 30, Appl
	19	34.4	5.7	246240	2	US-08-724-394A-20	Sequence 20, Appl
	20	34.4	5.7	246240	2	US-08-724-394A-21	Sequence 21, Appl
	21	34.4	5.7	246240	2	US-08-724-394A-22	Sequence 22, Appl
c	22	34	5.6	2218	2	US-08-421-044-1	Sequence 1, Appli
c	23	32.8	5.4	8353	3	US-08-611-587-1	Sequence 1, Appli
	24	31.8	5.2	1524	4	US-08-840-767-3	Sequence 3, Appli
c	25	31.2	5.1	5427	3	US-09-009-913-2	Sequence 2, Appli
c	26	31.2	5.1	5510	3	US-09-009-913-3	Sequence 3, Appli
c	27	31.2	5.1	5667	3	US-09-009-913-4	Sequence 4, Appli
	28	31.2	5.1	176373	3	US-09-128-155-17	Sequence 17, Appl
c	29	31	5.1	1082	2	US-08-909-965C-12	Sequence 12, Appl
c	30	30.6	5.0	1425	1	US-08-488-672-4	Sequence 4, Appli
c	31	30.6	5.0	6395	2	US-08-687-559-2	Sequence 2, Appli
c	32	30.6	5.0	6425	3	US-09-259-741-3	Sequence 3, Appli
c	33	30.6	5.0	6425	3	US-09-037-751-3	Sequence 3, Appli
c	34	30.6	5.0	6425	4	US-09-466-422-3	Sequence 3, Appli
c	35	30.6	5.0	6446	3	US-09-259-741-5	Sequence 5, Appli
c	36	30.6	5.0	6446	3	US-09-037-751-5	Sequence 5, Appli
c	37	30.6	5.0	6446	4	US-09-466-422-5	Sequence 5, Appli
c	38	30.6	5.0	6475	3	US-09-259-741-4	Sequence 4, Appli
c	39	30.6	5.0	6475	3	US-09-037-751-4	Sequence 4, Appli
c	40	30.6	5.0	6475	4	US-09-466-422-4	Sequence 4, Appli
c	41	30.6	5.0	7926	4	US-09-500-554-1	Sequence 1, Appli
c	42	30	5.0	1002	4	US-08-943-731-97	Sequence 97, Appl
	43	30	5.0	2766	3	US-08-936-135-3	Sequence 3, Appli
	44	30	5.0	3652	3	US-08-936-135-5	Sequence 5, Appli
c	45	30	5.0	38682	4	US-08-943-731-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-078-294-4
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 80246
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match 13.6%; Score 82.4; DB 4; Length 80246;
Best Local Similarity 63.8%; Pred. No. 4.2e-17;
Matches 125; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy      72 tgggtggggaacatgctactggcatctaatgcatagagggcagtaatgctgctaaacatc 131
           || |||||     ||||||||||||||| ||  | |||  |||  |||||| ||||||||
Db   66530 tgagtgggtgggttgctactggcatctagtgggtggagaccagagatgctgttaaacatc 66589

Qy     132 tttcaacgcacaggacagagccccacaaaagagaattatctagccccaaatgtccataac 191
             ||| |||||||||||  ||| |||| | ||||||||| |||||||| || |||
Db   66590 ccgcaaagcacaggacagtccccgacaacaaagaattatctggccccaaatatcaatagt 66649

Qy     192 actgctgttgagaaaacctaccgcaggatcttactgggcttcataggtaagcttgccttt 251
             |    ||||||||| |  |    ||  || |  |    |  |       |    |  ||
Db   66650 gccaaagttgagaaacctcattctagcttccttttcccttctacgttctaatcaactgtt 66709

Qy     252 gttctggcttctgtag 267
           |||||  |  |  |||
Db   66710 gttctttcagcattag 66725
```

RESULT 2
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.

```
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;    LENGTH: 80595
;    TYPE: DNA
;    ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3


  Query Match             13.6%;  Score 82.4;  DB 4;  Length 80595;
  Best Local Similarity   63.8%;  Pred. No. 4.2e-17;
  Matches  125;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy      72 tgggtggggaacatgctactggcatctaatgcatagagggcagtaatgctgctaaacatc 131
           || |||||     ||||||||||||||| ||  | |||  |||  |||||| ||||||||
Db   66788 tgagtgggtgggttgctactggcatctagtgggtggagaccagagatgctgttaaacatc 66847

Qy     132 tttcaacgcacaggacagagccccacaaaagagaattatctagccccaaatgtccataac 191
               ||| ||||||||||  ||| |||| | |||||||||| |||||||| || |||
Db   66848 ccgcaaagcacaggacagtccccgacaacaaagaattatctggccccaaatatcaatagt 66907

Qy     192 actgctgttgagaaaacctaccgcaggatcttactgggcttcataggtaagcttgccttt 251
           |    |||||||||  |  |    ||  || |  |    |  |      |     |  ||
Db -66908 gccaaagttgagaaacctcattctagcttcctttccctttctacgttctaatcaactgtt 66967

Qy     252 gttctggcttctgtag 267
           |||||  |  |  |||
Db   66968 gttctttcagcattag 66983


RESULT   3
US-09-078-294-13/c
; Sequence 13, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;    LENGTH: 7505
;    TYPE: DNA
;    ORGANISM: BAC-F2 contig 9
US-09-078-294-13


  Query Match             10.7%;  Score 64.8;  DB 4;  Length 7505;
  Best Local Similarity   72.1%;  Pred. No. 1e-11;
  Matches   98;  Conservative    0;  Mismatches   37;  Indels    1;  Gaps    1;

Qy      76 tggggaacatgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttc 135
           | |||||  |||||| | ||||||||   ||||||  ||  ||  || ||||  || | |
Db    4460 TTGGGAATTTGCTACAGACATCTAATAATTAGAGGCTAGGGATATTGTTAAATGTCCTAC 4401

Qy     136 aacgcacaggacagagccccacaaaagagaattatctagccccaaatgtccataacactg 195
           |  ||||| |||||| || || || |||||||| | |||||| |||||| ||| |   ||
Db    4400 AGTGCACATGACAGACCCACATAACAGAGAATTGTATAGCCCAAAATGT-CATTAGTGTG 4342

Qy     196 ctgttgagaaaacctа 211
           ||||||||||  ||
Db    4341 ACGTTGAGAAAATATA 4326


RESULT   4
US-08-211-430-1/c
; Sequence 1, Application US/08211430
; Patent No. 5763166
; GENERAL INFORMATION:
;  APPLICANT:
;    TITLE OF INVENTION:  NUCLEIC SEQUENCE OF THE GENE ASSOCIATED WITH
;    TITLE OF INVENTION:  X CHROMOSOME LINKED KALLMANN SYNDROME,CORRESPONDING
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES,DIAGNOSTIC APPLICATIONS.
;    NUMBER OF SEQUENCES:  32
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/211,430
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6314 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      CELL LINE:  foetal brain cells
US-08-211-430-1


  Query Match             10.0%;  Score 60.4;  DB 1;  Length 6314;
  Best Local Similarity   62.9%;  Pred. No. 3e-10;
  Matches  127;  Conservative    0;  Mismatches   71;  Indels    4;  Gaps    2;

Qy      15 tacccccatcccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctgg 74
           ||||  |   | || |     | |||||  ||  |||| | ||| | |      || ||
Db    5428 TACCATCCCTCAAGGGGTAGCTGGCAATATCTGGAGACATTTTTGGTTGTCTCAACTTG- 5368

Qy      75 gtggggaacatgctactggcatctaatgcatagagggcagtaatgctgctaaacatcttt 134
             |||||   ||| ||||||||||  ||   |||  |     ||||| ||||| | | |
Db    5369 --GGGGAGGGTGCAACTGGCATCTGGTGAGGAGAAAGTGTGGATGCTACTAAA-AACCTA 5313

Qy     135 caacgcacaggacagagccccacaaaagagaattatctagccccaaatgtccataacact 194
           ||  ||||| |||| | |||||||| |||||| |||| |||||||| ||| |||    |
Db    5312 CAGTGCACAAGACAAACCCCCACAACAGAGAACTATCCAGCCCCAAAAGTCAATAGTGTT 5253

Qy     195 gctgttgagaaaacctaccgca 216
           |  |||||||  | |  | | |
Db    5252 GAGGTTGAGATGAACCTCTGTA 5231


RESULT   5
US-09-157-177-120
; Sequence 120, Application US/09157177
; Patent No. 6090558
; GENERAL INFORMATION:
;  APPLICANT: Butler, John M.
;  APPLICANT: Li, Jia
;  APPLICANT: Monforte, Joseph A.
;  APPLICANT: Becker, Christopher H.
;  TITLE OF INVENTION: DNA TYPING BY MASS SPECTROMETRY WITH  POLYMORPHIC DNA
;  TITLE OF INVENTION: REPEAT MARKERS
;  FILE REFERENCE: GETR:017/GETR017P
;  CURRENT APPLICATION NUMBER: US/09/157,177
;  CURRENT FILING DATE: 1998-09-18
;  EARLIER APPLICATION NUMBER: 60/059,415
;  EARLIER FILING DATE: 1997-09-19
;  NUMBER OF SEQ ID NOS: 135
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 120
;    LENGTH: 279
;    TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
US-09-157-177-120


  Query Match             7.7%;  Score 46.8;  DB 3;  Length 279;
  Best Local Similarity   69.0%;  Pred. No. 2e-06;
  Matches   80;  Conservative    0;  Mismatches   32;  Indels    4;  Gaps    1;

Qy      48 aagacgtgttttgattgtcacacctgggtggggaacatgctactggcatctaatgcatag 107
           || |  | ||||| ||||||||  ||||  |     ||| ||||||||||  ||  | |
Db     165 aacaatttttttggttgtcacaaatgggatg----aatgttactggcatctggtgggtgg 220

Qy     108 agggcagtaatgctgctaaacatctttcaacgcacaggacagagccccacaaaaga 163
           ||  |||  |||||||| |||| | | ||  ||||| |||||| || |   |||||
Db     221 agcccagagatgctgctcaacaccctacagtgcacaagacagacccaccacaaaga 276


RESULT   6
US-09-276-531-34
; Sequence 34, Application US/09276531
; Patent No. 6183968
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Reddy, Roopa
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Baughn, Mariah R.
;    TITLE OF INVENTION:  COMPOSITION FOR THE DETECTION OF GENES ENCODING
;    TITLE OF INVENTION:  RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION
;    NUMBER OF SEQUENCES:  134
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/276,531
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/079,677
;      FILING DATE:  March 27, 1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lynn E. Murry, Ph.D.
;      REGISTRATION NUMBER:  42,918
;      REFERENCE/DOCKET NUMBER:  PA-0008 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1750 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  BRAITUT01
;      CLONE:  746982
US-09-276-531-34


  Query Match             6.8%;  Score 41;  DB 4;  Length 1750;
  Best Local Similarity   58.7%;  Pred. No. 0.00057;
  Matches  111;  Conservative    0;  Mismatches   70;  Indels    8;  Gaps    2;

Qy      22 atcccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctgggtgggga 81
           ||  || |    |||||  | | || |        ||  | ||||||||  |||  ||
Db    1258 ATTTCACTAGGCATTTGGCAATTCTTAGAGACATTTCCGGTTGTCACAATTGGAGGG--- 1314

Qy      82 acatgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacgca 141
           || ||| | ||||  | ||  |||||| |||  ||| |||||||  |  ||  ||  ||
Db    1315 --ATACTAGTAGCATGAATTGGGTAGAGGCCAGGGATGTTGCTAAATAGATTATAATACA 1372

Qy     142 caggacagagccccacaaaagagaattatctagccccaaatgtccataacactgctgttg 201
           || |  |     || |||  |||||||  |   ||| |||||||   |   ||  | ||
Db    1373 CAAGGAAAGCAGCCTCAA---AGAATTACCCCTCCCAAAATGTCAGAAGTGCTAAGGCTG 1429

Qy     202 agaaaacct 210
           ||||| |||
Db    1430 AGAAACCCT 1438


RESULT   7
US-08-646-301A-1/c
; Sequence 1, Application US/08646301A
; Patent No. 6194211
; GENERAL INFORMATION:
;  APPLICANT:  Richards, Cynthia Ann
;  APPLICANT:  Huber, Brian E.
;  TITLE OF INVENTION:  Transcriptional Regulatory Sequence of Carcinoembryonic
; Patent No. 6194211
;  TITLE OF INVENTION:  Antigen for Expression Targeting
;  FILE REFERENCE:  PB1508USW
;  CURRENT APPLICATION NUMBER: US/08/646,301A
;  CURRENT FILING DATE:  1996-05-16
;  NUMBER OF SEQ ID NOS:  25
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 1
;    LENGTH: 11288
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-08-646-301A-1


  Query Match             6.4%;  Score 38.6;  DB 4;  Length 11288;
  Best Local Similarity   62.6%;  Pred. No. 0.011;
  Matches   77;  Conservative    0;  Mismatches   44;  Indels    2;  Gaps    1;

Qy      25 ccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctgggtggggaaca 84
           ||||  | | ||||| ||    ||  |||| |  || | |  |    | ||||  ||||
Db    4309 CCAGGGGACATTTGGAAAGTCTGGAGACATTGTTGGTTGTTAGAGCTTGGGGAGGGA--G 4252

Qy      85 tgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacgcacag 144
           | || ||||||| || ||  |||||  |    ||| |||| |||| | | ||||  ||||
Db    4251 TCCTGCTGGCATTTAGTGGGTAGAGTCCCCAGATGGTGCTGAACACCCTACAACTTACAG 4192

Qy     145 gac 147
           ||
Db    4191 AAC 4189


RESULT   8
US-08-481-968A-4/c
; Sequence 4, Application US/08481968A
; Patent No. 6300490
; GENERAL INFORMATION:
;  APPLICANT:  Huber, Brian
;  APPLICANT:  Richards, Cynthia
;  TITLE OF INVENTION:  Molecular Constructs Comprising a Carcinoembryonic Antige
;  TITLE OF INVENTION:  Transcriptional Regulatory Region
;  FILE REFERENCE:  PB1087US4
;  CURRENT APPLICATION NUMBER: US/08/481,968A
;  CURRENT FILING DATE:  1998-06-07
;  NUMBER OF SEQ ID NOS:  36
;  SOFTWARE:  PatentIn version 3.0
; SEQ ID NO 4
```

```
;   LENGTH: 11288
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-08-481-968A-4


  Query Match             6.4%;  Score 38.6;  DB 4;  Length 11288;
  Best Local Similarity   62.6%;  Pred. No. 0.011;
  Matches   77;  Conservative    0;  Mismatches   44;  Indels    2;  Gaps    1;

Qy      25 ccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctgggtggggaaca 84
           |||| | | ||||| ||   ||  |||| |  || | |  |    | ||||  ||||
Db    4309 CCAGGGGACATTTGGAAAGTCTGGAGACATTGTTGGTTGTTAGAGCTTGGGGAGGGA--G 4252

Qy      85 tgctactggcatctaatgcatagagggcagtaatgctgctaaacatctttcaacgcacag 144
           | || ||||||| || ||  |||||  |    ||| |||| |||| | | ||||  ||||
Db    4251 TCCTGCTGGCATTTAGTGGGTAGAGTCCCCAGATGGTGCTGAACACCCTACAACTTACAG 4192

Qy     145 gac 147
           ||
Db    4191 AAC 4189


RESULT   9
US-08-910-733-17/c
; Sequence 17, Application US/08910733
; Patent No. 5837495
;  GENERAL INFORMATION:
;    APPLICANT:  COLOTTA, Francesco
;    APPLICANT:  MUZIO, Marta
;    APPLICANT:  MANTOVANI, Alberto
;    TITLE OF INVENTION:  INTERLEUKIN-1 ANTAGONIST, DNA ENCODING SAME,
;    TITLE OF INVENTION:  AND ANTIBODIES THERETO
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/910,733
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/476,860
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  IT MI 94 A 002097
;      FILING DATE:  13-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  YUN, Allen C.
;      REGISTRATION NUMBER:  37,971
;      REFERENCE/DOCKET NUMBER:  COLOTTA=1A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3176 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US 08-910-733-17


  Query Match             6.0%;  Score 36.6;  DB 2;  Length 3176;
  Best Local Similarity   65.1%;  Pred. No. 0.026;
  Matches   54;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy     387 agaatgaccaaccatcctggttggcctgggactgtcctagttttagcattgaaagtttca 446
           ||| |||||||| | | ||||| | |  ||||| || |   ||||  | ||||||||  |
Db    2648 AGAGTGACCAACTACCTTGGTTTGTCCAGGACTTTCTTGACTTTATTACTGAAAGTTCTA 2589

Qy     447 ggttccaggaaagccctcaggcc 469
               ||| ||||||| ||
Db    2588 CACCTTGCCAAACCCCTCAGTCC 2566


RESULT  10
US-08-910-884-17/c
; Sequence 17, Application US/08910884
; Patent No. 5981713
;  GENERAL INFORMATION:
;    APPLICANT:  COLOTTA, Francesco
;    APPLICANT:  MUZIO, Marta
;    APPLICANT:  MANTOVANI, Alberto
;    TITLE OF INVENTION:  INTERLEUKIN-1 ANTAGONIST, DNA ENCODING SAME,
;    TITLE OF INVENTION:  AND ANTIBODIES THERETO
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/910,884
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/476,860
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  IT MI 94 A 002097
;      FILING DATE:  13-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  YUN, Allen C.
;      REGISTRATION NUMBER:  37,971
;      REFERENCE/DOCKET NUMBER:  COLOTTA=1B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3176 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-910-884-17


  Query Match             6.0%;  Score 36.6;  DB 2;  Length 3176;
  Best Local Similarity   65.1%;  Pred. No. 0.026;
  Matches   54;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy     387 agaatgaccaaccatcctggttggcctgggactgtcctagttttagcattgaaagtttca 446
           ||| |||||||| | | ||||| | |  ||||| || |   ||||  | ||||||||  |
Db    2648 AGAGTGACCAACTACCTTGGTTTGTCCAGGACTTTCTTGACTTTATTACTGAAAGTTCTA 2589

Qy     447 ggttccaggaaagccctcaggcc 469
               ||| ||||||| ||
```

```
Db   2588 CACCTTGCCAAACCCCTCAGTCC 2566


RESULT 11
US-08-068-945A-1
; Sequence 1, Application US/08068945A
; Patent No. 5616483
;  GENERAL INFORMATION:
;    APPLICANT:  Bjursell, Gunnar
;    APPLICANT:  Carlsson, Peter
;    APPLICANT:  Enerback, Sven
;    APPLICANT:  Hansson, Lennart
;    APPLICANT:  Lidberg, Ulf
;    APPLICANT:  Nilsson, Jeanette
;    APPLICANT:  Tornell, Jan
;    TITLE OF INVENTION:  New DNA Sequences
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  White & Case
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2787
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/068,945A
;      FILING DATE:  27-MAY-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 9201809-2
;      FILING DATE:  11-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 9201826-6
;      FILING DATE:  12-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 9202088-2
;      FILING DATE:  03-JUL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 9300902-5
;      FILING DATE:  19-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sterner, Richard J.
;      REGISTRATION NUMBER:  35,372
;      REFERENCE/DOCKET NUMBER:  1103326-052
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)819-8783
;      TELEFAX:  (212)354-8113
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11531 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Mammary gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1653..1727, 4071..4221, 4307..4429, 4707
;      LOCATION:  ..4904, 6193..6323, 6501..6608, 6751..6868, 8335
;      LOCATION:  ..8521, 8719..8922, 10124..10321, 10650..11394)
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  join(1722..1727, 4071..4221, 4307..4429, 4707
;      LOCATION:  ..4904, 6193..6323, 6501..6608, 6751..6868, 8335
;      LOCATION:  ..8521, 8719..8922, 10124..10321, 10650..11391)
;      OTHER INFORMATION:  /EC_number= 3.1.1.1
;      OTHER INFORMATION:  /product= "Bile Salt-Stimulated Lipase"
;    FEATURE:
;      NAME/KEY:  5'UTR
;      LOCATION:  1..1640
;    FEATURE:
;      NAME/KEY:  TATA_signal
;      LOCATION:  1611..1617
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1641..1727
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  4071..4221
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  4307..4429
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  4707..4904
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  6193..6323
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  6501..6608
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  6751..6868
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  8335..8521
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  8719..8922
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  10124..10321
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  10650..11490
;    FEATURE:
;      NAME/KEY:  3'UTR
;      LOCATION:  11491..11531
US-08-068-945A-1

  Query Match             5.9%;  Score 35.6;  DB 1;  Length 11531;
  Best Local Similarity   61.4%;  Pred. No. 0.12;
  Matches  108;  Conservative    0;  Mismatches   64;  Indels    4;  Gaps    3;

Qy      14 ttacccccatcccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctg 73
           | || | || ||  | ||    | |||   |||  |||| |  | ||||||||||| |||
Db    8055 TGACTCTCAGCCTGGGGACGTGTGGCAGGGCCTGGAGACAT-CTGTGATTGTCACAGCTG 8113

Qy      74 ggtggggaacatgctactggcatctaatgcatagagggcagtaatgctgctaaacatctt 133
             ||  |||    |||| |||||| ||  ||  | |||| | |  ||||| |||||||||| |
Db    8114 GGGAGGG--GGTGCTCCTGGCACCTCGTGGGTCGAGGCCGGGGATGCT-CTAAACATCCT 8170

Qy     134 tcaacgcacaggacagagccccacaaaagagaattatctagccccaaatgtccata 189
             ||  ||||||||     |          |   | |  ||||||| ||||||||
Db    8171 ACAGGGCACAGGATGCCCCTGATGGTGCAGAATCAACCCTGCCCCAAGTGTCCATA 8226


RESULT 12
US-08-442-806-1
; Sequence 1, Application US/08442806
; Patent No. 5716817
;  GENERAL INFORMATION:
;    APPLICANT:  Bjursell, Gunnar
;    APPLICANT:  Carlsson, Peter
;    APPLICANT:  Enerback, Sven
;    APPLICANT:  Hansson, Lennart
;    APPLICANT:  Lidberg, Ulf
```

```
;    APPLICANT:  Nilsson, Jeanette
;    APPLICANT:  Tornell, Jan
;    TITLE OF INVENTION:  Genomic DNA Sequences
;    TITLE OF INVENTION:  Encoding Human BSSL/CEL
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  White & Case
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2787
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/442,806
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/068,945
;      FILING DATE:  27-MAY-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 9201809-2
;      FILING DATE:  11-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 9201826-6
;      FILING DATE:  12-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 9202088-2
;      FILING DATE:  03-JUL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 9300902-5
;      FILING DATE:  19-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sterner, Richard J.
;      REGISTRATION NUMBER:  35,372
;      REFERENCE/DOCKET NUMBER:  1103326-052
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)819-8783
;      TELEFAX:  (212)354-8113
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11531 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Mammary gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1653..1727, 4071..4221, 4307..4429, 4707
;      LOCATION:  ..4904, 6193..6323, 6501..6608, 6751..6868, 8335
;      LOCATION:  ..8521, 8719..8922, 10124..10321, 10650..11394)
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  join(1722..1727, 4071..4221, 4307..4429, 4707
;      LOCATION:  ..4904, 6193..6323, 6501..6608, 6751..6868, 8335
;      LOCATION:  ..8521, 8719..8922, 10124..10321, 10650..11391)
;      OTHER INFORMATION:  /EC_number= 3.1.1.1
;      OTHER INFORMATION:  /product= "Bile Salt-Stimulated Lipase"
;    FEATURE:
;      NAME/KEY:  5'UTR
;      LOCATION:  1..1640
;    FEATURE:
;      NAME/KEY:  TATA_signal
;      LOCATION:  1611..1617
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1641..1727
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  4071..4221
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  4307..4429
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  4707..4904
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  6193..6323
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  6501..6608
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  6751..6868
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  8335..8521
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  8719..8922
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  10124..10321
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  10650..11490
;    FEATURE:
;      NAME/KEY:  3'UTR
;      LOCATION:  11491..11531
US-08-442-806-1


  Query Match             5.9%;  Score 35.6;  DB 1;  Length 11531;
  Best Local Similarity   61.4%;  Pred. No. 0.12;
  Matches  108;  Conservative    0;  Mismatches   64;  Indels    4;  Gaps    3;

Qy      14 ttacccccatcccagtgaatatttgcaattcctaaagacgtgttttgattgtcacacctg 73
           | || | || ||  | ||    | |||   ||| |||| |  | ||||||||||| |||
Db    8055 TGACTCTCAGCCTGGGGACGTGTGGCAGGGCCTGGAGACAT-CTGTGATTGTCACAGCTG 8113

Qy      74 ggtggggaacatgctactggcatctaatgcatagagggcagtaatgctgctaaacatctt 133
           ||  |||    |||| |||||| ||  ||  | |||| | |  ||||| |||||||||| |
Db    8114 GGGAGGG--GGTGCTCCTGGCACCTCGTGGGTCGAGGCCGGGGATGCT-CTAAACATCCT 8170

Qy     134 tcaacgcacaggacagagccccacaaaagagaattatctagccccaaatgtccata 189
            ||  |||||||||     |          |   | |  ||||||| ||||||||
Db    8171 ACAGGGCACAGGATGCCCCTGATGGTGCAGAATCAACCCTGCCCCAAGTGTCCATA 8226


RESULT  13
US-09-429-323-3
; Sequence 3, Application US/09429323A
; Patent No. 6140126
; Patent No. 6140126 6140123
; GENERAL INFORMATION:
;  APPLICANT: C. Frank Bennett
;  APPLICANT: Lex M. Cowsert
;  TITLE OF INVENTION: ANTISENSE MODULATION OF Y-BOX BINDING PROTEIN 1 EXPRESSION
;  FILE REFERENCE: RTS-0092
;  CURRENT APPLICATION NUMBER: US/09/429,323A
;  CURRENT FILING DATE: 1999-10-26
;  NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
;   LENGTH: 1481
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
```

```
;   NAME/KEY: CDS
;   LOCATION: (127)..(1080)
US-09-429-323-3


  Query Match          5.8%;  Score 35.4;  DB 3;  Length 1481;
  Best Local Similarity  47.9%;  Pred. No. 0.042;
  Matches  102;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy      70 cctgggtggggaacatgctactggcatctaatgcatagagggcagtaatgctgctaaaca 129
           || | | ||  || | | |  | ||| | || |  | | |  ||||||    | | ||
Db     269 ccggcggggacaagaaggtcatcgcaacgaaggttttgggaacagtaaaatggttcaatg 328

Qy     130 tctttcaacgcacaggacagagccccacaaaagagaattatctagccccaaatgtccata 189
           |     | |    ||    | |  ||  || || |   |   ||    |  ||| ||
Db     329 taaggaacggatatggtttcatcaacaggaatgacaccaaggaagatgtatttgtacacc 388

Qy     190 acactgctgttgagaaaacctaccgcaggatcttactgggcttcataggtaagcttgcct 249
           | |||||  |  |||| |   ||| |||||  |  ||  ||    ||||  |    |
Db     389 agactgccataaagaagaataaccccaggaagtaccttcgcagtgtaggagatggagaga 448

Qy     250 ttgttctggcttctgtagatatataaaataaag 282
           |||   |  |  ||| | |  |  | | || |
Db     449 ctgtggagtttgatgttgttgaaggagaaaagg 481


RESULT  14
US-07-688-352C-31
; Sequence 31, Application US/07688352C
; Patent No. 5527896
;  GENERAL INFORMATION:
;    APPLICANT:  Wigler, Michael H.
;    APPLICANT:  Colicelli, John J.
;    TITLE OF INVENTION:  Cloning by Complementation and Related
;    TITLE OF INVENTION:  Processes
;    NUMBER OF SEQUENCES:  57
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  Two First National Plaza, 20 South Clark
;      STREET:  Street
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/688,352C
;      FILING DATE:  19910419
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/511,715
;      FILING DATE:  20-APR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Borun, Michael F.
;      REGISTRATION NUMBER:  25447
;      REFERENCE/DOCKET NUMBER:  27805/30197
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 346-5750
;      TELEFAX:  (312) 984-9740
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3073 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..1111
US-07-688-352C-31


  Query Match          5.8%;  Score 35.4;  DB 1;  Length 3073;
  Best Local Similarity  47.9%;  Pred. No. 0.064;
  Matches  102;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy      70 cctgggtggggaacatgctactggcatctaatgcatagagggcagtaatgctgctaaaca 129
           || | | ||  || | | |  | ||| | || |  | | |  ||||||    | | ||
Db     280 CCGGCGGGGACAAGAAGGTCATCGCAACGAAGGTTTTGGGAACAGTAAAATGGTTCAATG 339

Qy     130 tctttcaacgcacaggacagagccccacaaaagagaattatctagccccaaatgtccata 189
           |     | |    ||    | |  ||  || || |   |   ||    |  ||| ||
Db     340 TAAGGAACGGATATGGTTTCATCAACAGGAATGACACCAAGGAAGATGTATTTGTACACC 399

Qy     190 acactgctgttgagaaaacctaccgcaggatcttactgggcttcataggtaagcttgcct 249
           | |||||  |  |||| |   ||| |||||  |  ||  ||    ||||  |    |
Db     400 AGACTGCCATAAAGAAGAATAACCCCAGGAAGTACCTTCGCAGTGTAGGAGATGGAGAGA 459

Qy     250 ttgttctggcttctgtagatatataaaataaag 282
           |||   |  |  ||| | |  |  | | || |
Db     460 CTGTGGAGTTTGATGTTGTTGAAGGAGAAAAGG 492


RESULT  15
US-08-474-379C-31
; Sequence 31, Application US/08474379C
; Patent No. 5977305
;  GENERAL INFORMATION:
;    APPLICANT:  Wigler, Michael H.
;    APPLICANT:  Colicelli, John J.
;    TITLE OF INVENTION:  CLONING BY COMPLEMENTATION AND RELATED
;    TITLE OF INVENTION:  PROCESSES
;    NUMBER OF SEQUENCES:  88
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  233 South Wacker Drive/6300 Sears Tower
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,379C
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/511,715
;      FILING DATE:  20-APR-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/206,188
;      FILING DATE:  01-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/688,352
;      FILING DATE:  19-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  27866/32771
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3073 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..1109
US-08-474-379C-31

Query Match 5.8%; Score 35.4; DB 2; Length 3073;
Best Local Similarity 47.9%; Pred. No. 0.064;
Matches 102; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```
Qy      70 cctgggtggggaacatgctactggcatctaatgcatagagggcagtaatgctgctaaaca 129
           || | | ||  || | | |  | ||| | || |  | | |  ||||||    | | ||
Db     280 CCGGCGGGGACAAGAAGGTCATCGCAACGAAGGTTTTGGGAACAGTAAAATGGTTCAATG 339

Qy     130 tctttcaacgcacaggacagagccccacaaaagagaattatctagccccaaatgtccata 189
           |     | |    ||    | |  ||  || || |   |   ||    |  ||| ||
Db     340 TAAGGAACGGATATGGTTTCATCAACAGGAATGACACCAAGGAAGATGTATTTGTACACC 399

Qy     190 acactgctgttgagaaaacctaccgcaggatcttactgggcttcataggtaagcttgcct 249
           | |||||  |  |||| |   ||| |||||  |  ||  ||    |||| |    |
Db     400 AGACTGCCATAAAGAAGAATAACCCCAGGAAGTACCTTCGCAGTGTAGGAGATGGAGAGA 459

Qy     250 ttgttctggcttctgtagatatataaaataaag 282
           |||   |  |  ||| | |  |  | | || |
Db     460 CTGTGGAGTTTGATGTTGTTGAAGGAGAAAAGG 492
```

Search completed: November 5, 2001, 22:58:05
Job time: 2485 sec